

# Psychophysiology and the detection of deception: promises and perils

Citation for published version (APA):

Meijer, E. H. (2008). *Psychophysiology and the detection of deception: promises and perils*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht.  
<https://doi.org/10.26481/dis.20081107em>

## Document status and date:

Published: 01/01/2008

## DOI:

[10.26481/dis.20081107em](https://doi.org/10.26481/dis.20081107em)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

*Psychophysiology and the detection of deception:*

**Promises and perils**

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht, op  
gezag van de Rector Magnificus, prof. mr. G.P.M.F. Mols volgens het  
besluit van het College van Decanen, in het openbaar te verdedigen op  
vrijdag 7 november 2008 om 14.00 uur

door

Ewout Hendrik Meijer



**Promotor:**

Prof. dr. H.L.G.J. Merckelbach

**Copromotor:**

Dr. F.T.Y. Smulders

**Beoordelingscommissie:**

Prof. dr. P.J. van Koppen (voorzitter)

Dr. B.M. Jansma

Prof. dr. J.L. Kenemans (Universiteit Utrecht)

Prof. dr. C. de Ruiter

Prof. dr. A. Vrij (University of Portsmouth, United Kingdom)

The research presented in this dissertation was funded by the *Wetenschappelijk Onderzoeks en Documentatiecentrum* (WODC) of the Dutch Ministry of Justice.

Cover illustration: Der Lügendetektor, 1990, Object. By Wolfgang Nieblich

Druk: Datawyse / Universitaire Pers Maastricht

ISBN: 978-90-5278-763-3

© Ewout H. Meijer, Maastricht 2008

<b>Prelude .....</b>	<b>7</b>
<b>Part I.....</b>	<b>8</b>
<b>Chapter 1: Lie detectors and the law: The use of the polygraph in Europe .....</b>	<b>9</b>
1.1 <i>The polygraph</i> .....	11
1.2 <i>Accuracy</i> .....	14
1.3 <i>Legal systems</i> .....	15
1.4 <i>The polygraph in Europe</i> .....	16
<b>Chapter 2: Sex offender management using the polygraph: A critical review .....</b>	<b>21</b>
2.1 <i>Introduction</i> .....	22
2.2 <i>The polygraph and the detection of deception</i> .....	24
2.3 <i>Polygraph validity</i> .....	25
2.4 <i>The polygraph and sex offenders</i> .....	26
2.5 <i>Disclosure through polygraph testing</i> .....	29
2.6 <i>Bogus pipeline</i> .....	30
2.7 <i>Reducing recidivism</i> .....	32
2.8 <i>Concluding remarks</i> .....	33
<b>Part II.....</b>	<b>34</b>
<b>Chapter 3: Concealed information .....</b>	<b>35</b>
3.1 <i>Introduction</i> .....	36
3.2 <i>Skin Conductance Response</i> .....	37
3.3 <i>P300</i> .....	38
3.4 <i>SVT</i> .....	39
3.5 <i>Outline</i> .....	40
<b>Chapter 4: The P300 is Sensitive to Concealed Face Recognition.....</b>	<b>43</b>
4.1 <i>Introduction</i> .....	44
4.2 <i>Experiment 1</i> .....	46
4.3 <i>Experiment 2</i> .....	50
4.4 <i>General discussion</i> .....	52

<b>Chapter 5: Psychopathic traits and the detection of concealed information: Event related potential and skin conductance combined .....</b>	<b>55</b>
5.1 Introduction .....	56
5.2 Method .....	58
5.3 Results .....	62
5.4 Discussion .....	66
<b>Chapter 6: Combining Skin Conductance and Forced Choice in the Detection of Concealed Information.....</b>	<b>69</b>
6.1 Introduction .....	70
6.2 Experiment 1 .....	74
6.3 Experiment 2 .....	78
6.4 General discussion.....	81
<b>General discussion .....</b>	<b>91</b>
Mechanisms underlying the P300 in a CIT.....	93
Towards application of the CIT.....	98
fMRI and the detection of deception .....	99
Voice analysis and SCAN.....	101
Future Research.....	104
Concluding remarks.....	106
<b>References.....</b>	<b>107</b>
<b>Summaries.....</b>	<b>123</b>
English .....	123
Nederlands .....	125
<b>Curriculum vitae.....</b>	<b>127</b>
<b>Publications .....</b>	<b>128</b>
<b>Dankwoord.....</b>	<b>130</b>



### **Prelude**

The number of situations in which the veracity of a statement is debated is numerous. Recent mass media covered cases include Rabobank cyclist Rasmussen: was he lying about his 2007 pre-tour whereabouts? And Joran van der Sloot: was he lying about his role in the disappearance of Natalee Holloway? And what about the United States government: were they lying about their knowledge of the presence of weapons of mass destruction in Iraq? Of course, daily life examples are also widely available. Is this job applicant being truthful? Did my partner cheat on me? Is this parolee adhering to his parole conditions? Has this child been abused? Given the limited capability of even trained detectives in distinguishing truth from lies (Stroemwall, Granhag, & Hartwig, 2004; Vrij, 2008), it is not surprising that resorting to technological aids has been popular throughout history. Of these technological aids, the polygraph has by far received the most attention.

Still, polygraph tests are highly controversial. They are laughed off as pseudoscience by some (Hines, 2003), while portrayed as highly scientific by others (e.g., Undeutsch, 1996). And as the debate surrounding polygraph tests advanced into the 21<sup>st</sup> century, it also came closer: polygraph tests are becoming increasingly popular in Europe, in a variety of settings. This dissertation critically evaluates different techniques that are used for the detection of deception. It consists of two parts. The first part addresses the widely used Control Question polygraph Test (CQT). Chapter 1 explains this type of polygraph test, and describes how it is employed in judicial procedures in different countries around Europe. Chapter 2 critically discusses a specific application of the CQT that is increasingly becoming popular, namely the periodical testing of sex offenders. The second part of this dissertation focuses on a different forensic application of the polygraph, namely the Concealed Information polygraph Test (CIT). In three chapters, we discuss different measures and combination of measures aimed at improving the detection efficiency of the CIT. After explaining the nature of the CIT in Chapter 3, Chapter 4 addresses the use of the P300 event-related brain potential in the detection of concealed face recognition. Chapter 5 tests the combination of skin conductance and P300 in a CIT. In Chapter 6, we describe the combination of skin conductance and a forced choice measure. In the final chapter, the work is summarized and both theoretical and practical implications are discussed. Attention is given to recent developments in the domain of lie detection, and directions for future research are discussed.

# Part I

## Chapter 1: Lie detectors and the law: The use of the polygraph in Europe

Adapted from:

Meijer, E. H., & van Koppen, P. J. (2008). Lie detectors and the law: The use of the polygraph in Europe. In D. Canter & R. Zukauskienė (Eds.), *Psychology and law: Bridging the gap* (pp. 31-50). Aldershot, UK: Ashgate Publishing.

### **Abstract**

The Control Question polygraph Test (CQT) is becoming increasingly popular in Europe. In this chapter, we discuss the nature of this lie detection test, and review how far it has penetrated into the legal arena of different European countries. The CQT is used in police investigations on a relatively large scale in countries like Belgium and Finland. Courts around Europe seem to be aware of the controversial nature of this polygraph technique, and have generally rejected it as evidence. Due to the free evidence nature of many countries governing evidence, however, this has never been a general ban. Moreover, tests can be used in criminal investigations, without ever reaching the courtroom.



## CHAPTER 1

*‘Whenever I lecture on polygraphic interrogation, I like to begin by asking the audience to indicate by show of hands how many would agree to take a lie detector test – and how many would refuse - in some plausible, hypothetical situation that I briefly describe. Among American audiences, a substantial majority acknowledge that they would take the test if they were innocent of wrongdoing and wished to prove their truthfulness. Many Europeans, in contrast, have never heard of any such device. The lie detector is almost exclusively an American artifact, and in the United States, it seems firmly entrenched in popular mythology (Lykken, 1998, p. 5).’*

This is how, 10 years ago, psychologist David Lykken opened the second edition of his book on uses and abuses of the lie detector. But times have changed, and over the last decade, lie detectors, or polygraph tests as they are also known, have been introduced in various settings around Europe. A striking example is the situation in Belgium. In the nineties, the Belgian federal police found themselves confronted with a number of major criminal investigations they had been unable to solve. In order to force a breakthrough, they decided to submit suspects to polygraph tests. Experienced polygraphers were flown in; from South Africa to test Flemish suspects, from Canada for Wallonian suspects. The introduction of the polygraph in Belgium turned out to be successful. Not necessarily because the cases were solved by the polygraph per se, but more so because a substantial proportion of suspects confessed during or after the test.<sup>1</sup> Following these successes, the Belgian Federal Police in Brussels had their own staff trained. Currently, three dedicated polygraphers test over 300 suspects per year.

The Belgian example illustrates that polygraph testing is no longer exclusively an American affair. Like Belgium, some other European countries have adopted the use of polygraph testing to some extent, while others have not. Meanwhile, the merits and demerits of polygraph tests are heavily debated. Arguments for this debate not only include scientific properties of the test, like accuracy, but also legal and ethical issues, for example, whether it’s alien to the right to remain silent. Legal systems in different European countries weigh these issues differently, resulting in different legal restraints on the use of polygraph test outcomes.

---

<sup>1</sup> For example: in 2003 the Belgian Police used the polygraph in 99 investigations, testing a total of 251 individuals. One-hundred-and-forty of these tests were in Dutch, and of these 140 tests 30 had a ‘deception indicated’ outcome. Five suspects confessed directly to the polygrapher when confronted with this outcome. Another 5 confessed at a later stage. Also, two suspects confessed before the test actually started.

## LIE DETECTORS AND THE LAW

In this chapter, we will discuss the use of the polygraph in Europe, and how legal systems in various countries deal with the issues surrounding it. We do not pretend to give an exhaustive review. Such review would be nearly impossible, because polygraph testing is partially in the hands of secret services and the military, and thus confidential. Even information that is not confidential is rarely published and often inaccessible. Therefore, only applications of the polygraph for which information is publicly available will be discussed. For a better understanding of the issues affecting legal status, we will first discuss the basic principles of polygraph testing. Because the practical application of the polygraph in Europe is limited to the use of the Control Question polygraph Test, we will only describe this type of test.<sup>2</sup> After that, we will discuss how far polygraph testing has penetrated into the legal arena of several European countries.

### *1.1 The polygraph*

The words ‘polygraph’ and ‘lie detector’ are often used synonymously. Technically, the term ‘polygraph’ refers only to the recording device that is used for registering different physiological parameters. Polygraphs that are specifically designed for lie detection tests used to be briefcase-sized machines that registered physiological signals from the subject by recording them with multiple pens on a lengthy roll of paper. Nowadays, they consist of a small amplifier/digitizer and a laptop recording the signals. The sensors attached to the subject are generally (1) two expendable bands positioned around the thorax measuring respiration, (2) two electrodes attached to the inside of the hand measuring palmar sweating and (3) an inflatable cuff positioned around the upper arm registering blood pressure. These physiological parameters co-vary with a number of psychological processes, including orienting and emotions like fear and stress. Although it is widely accepted that there is no unique physiological pattern associated with lying, the tracings can still be used to infer guilt or innocence, as will be discussed below.

The most widely used polygraph technique is known as the Control Question Test (CQT; Reid, 1947). During this type of test, the suspect is asked three types of questions, while connected to the recording device. The first type, the relevant questions, deal directly with the incident under investigation, for example “on the 25th of March, did you shoot John Doe?” The second type, the control questions, have a more generic nature, but also deal with undesirable behavior, for example “In the first 25 years of

---

<sup>2</sup> In this chapter, the terms lie detector, polygraph and CQT are used interchangeably.



## CHAPTER 1

your life, have you ever done anything illegal?” The third type, the irrelevant questions, are neutral, for example, “Is today Wednesday?” and are used as fillers. The physiological reactions to the relevant questions are compared to those accompanying the control questions. If the relevant questions elicit the largest responses, test outcome is ‘deception indicated’. When the control questions elicit the largest responses, test outcome is ‘no deception indicated’.<sup>3</sup>

The CQT is thus based upon the assumptions that a guilty suspect will show the strongest reactions to the relevant question, while an innocent suspect will react most strongly to the control question. At first glance, there seems to be some face validity to the former assumption. It is reasonably imaginable that a guilty suspect perceives the relevant questions as most threatening. But why would an innocent suspect show the largest response to the control questions? This assumption requires some explanation. A typical CQT starts with a lengthy pre-test interview. During this interview, the suspect is led to believe that lying is accompanied by involuntary changes in physiological activity. It is further explained that the polygraph registers these changes, and can thus determine whether the suspect is lying or not. Most importantly, however, the suspect is told that for the test outcome to be ‘no deception indicated’ all questions need to be answered truthfully. Meanwhile, the polygrapher emphasizes the importance of answering both the relevant and the control questions with ‘No’. This can be achieved, for example, by giving the suspect the impression that disclosure of previous illegal behavior will negatively influence the polygrapher’s opinion of the suspect, and may thus negatively influence the test outcome. Offe and Offe (2007) give a good example of how this is achieved in an experiment where the theft of a €50 voucher is the incident under investigation:

*I am going to ask you some questions to find out what your history concerning this matter looks like. I want to give you the reason, too, why I have to ask you such indiscrete questions. I want to find out whether one would consider you capable of an action such as removing a voucher for 50 € based on your history or not. In a nutshell, I want to know whether such an action, taking something of monetary value out of a closed room fits your personality profile or not.*

*These personal questions also have to be answered entirely truthfully. The more of these questions you can truthfully negate, the better it is for you, because then one can say that such an action does not fit your personality profile. If, however, you have to truthfully answer yes, then I will have to continue asking what the context was, so that I can get an impression of whether these*

---

<sup>3</sup> If no clear difference between the physiological responses to the different questions is observed, test outcome is ‘inconclusive’.

## LIE DETECTORS AND THE LAW

*were small and harmless delinquencies or whether there were some serious ones as well. Depending on what you tell me, it may begin to become imaginable that you may have done what we are talking about here as well.*

*The personal questions have nothing to do with whether you have taken the voucher or not. In order to make that clear from the beginning of every question, each personal question will start with: "Prior to 1999, ...," so that you will know right away, "this is about my past." For the result of the polygraph examination it is important, that you answer these questions truthfully as well.*

Assuming that at some points in their youth, everybody has overstepped the mark, the suspect will feel forced to give a deceptive answer to each control question. This should create a situation in which for an innocent suspect the control questions become most stressful. After all, he or she is forced to give a deceptive answer, and led to believe that this will be picked up by the polygraph, and will cause him or her to fail the entire test. As said, a guilty suspect is thought to be less worried about the control questions. For this person, the deceptive answer to the relevant questions poses the greatest threat of failing the test. This assumption underlying the CQT has also been referred to as the inference rule: "each examinee will focus his or her concern on the questions that present the greatest threat of failing the test" (Elaad, 2003, p. 38).

Still, the application of CQT remains controversial, with especially the assumption of innocent suspects showing the greatest concern to the control questions being problematic. After all, the stress inducing effect is not a feature of the control question itself, but rather a consequence of how its function is explained to the suspect. The extent to which the control questions actually induce stress depends largely on the skills of the polygrapher, and understanding of the suspect (Offe & Offe, 2007).<sup>4</sup>

---

<sup>4</sup> This means that, in order to be able to evaluate test quality, documentation of the entire test needs to be available, which is not always the case. The following is an example. In a civil case, we were asked to give our opinion on the quality of a CQT, and the weight that should be given to its outcome. The defendant underwent a polygraph test with a private commercial polygrapher from London. The documentation of the test consisted of a sworn affidavit and a videotape. In the sworn affidavit, the polygrapher claimed a nearly 100% success rate (the only exceptions being when a client cut short an investigation), stated the relevant questions as well as his conclusion: "I can state categorically that all your replies were truthful". In the videotape the polygrapher addressed the ex-wife of the accused: "There's no way that the accused is having any form of sexual contact with your children or with anybody else's children. He is not attracted to children. He is however very very much in love with you." When trying to get hold of more documentation, it turned out that there was no audio or video recording of any portion of the test itself, because this, according to the polygrapher, would have distracted the examinee. The videotape did contain fragments from before and after the



## CHAPTER 1

According to proponents of the CQT, a skilled polygrapher is capable of both formulating control questions and creating an atmosphere in which an innocent examinee is most worried about these. CQT critics argue, on the other hand, that this assumption has no grounding in psychological or psychophysiological research, nor is it convincing in its inner logic (see e.g., Ben-Shakhar, in press).

### **1.2 Accuracy**

Estimates of the error rates of the CQT vary. This variation is in part due to difficulties that characterize research on the method's accuracy. In a typical laboratory study, subjects are instructed to commit a mock crime and are subsequently tested with a CQT. Their results are compared to tested subjects who did not commit the mock crime. Problematic here is that the CQT is based on the detection of emotions like fear and stress. These are likely to differ between laboratory studies where an undergraduate psychology student – serving as a subject to gain study credits – is suspected of having committed a mock crime and real life cases where the defendant is suspected of murder and faces life imprisonment. Field studies have more ecological validity, but are plagued by other problems. Most importantly, they lack an objective criterion of guilt or innocence (i.e., ground truth). Whether the suspect was convicted afterwards is a suboptimal criterion, since a conviction may be, directly or indirectly, influenced by the results of the CQT. Especially admissions or confessions made by suspects when confronted with a polygraph test result in a sampling bias overestimating the validity of the CQT (Iacono, 1991).

Keeping these limitations in mind, mock crime studies come up with accuracy estimates that range from 74% to 80% for guilty participants, and from 63% to 66% for innocent participants (Ben Shakhar & Furedy, 1990; Kircher, Horowitz, & Raskin, 1988). The false positive ratio (i.e., innocent examinees classified as guilty) is estimated to be in 12% to 15% range, and the false negative ratio (i.e., guilty examinees classified as innocent) is 7% to 8%. Field study estimates of the accuracy of the CQT are slightly higher, ranging from 84% to 89% for guilty, and from 59% to 72% for innocent suspects, with estimated false positive rates of 12% to 23% and

---

test. Strikingly, this showed the examinee seated next to a running washing machine. Furthermore, the polygrapher had disposed of the graphs because he maintained a paperless office. We concluded that *'these materials (and the lack thereof), indicate either an amateurish test, or a deliberate attempt to make the test outcome unverifiable. In either case, no weight whatsoever should be given to the test outcome'*.

## LIE DETECTORS AND THE LAW

false negative rates of 1% to 13% (Ben Shakhar & Furedy, 1990; Raskin & Honts, 2002).<sup>5</sup> More recently, a panel of 14 leading American scientists reviewed the literature on the accuracy of the CQT. The 37 laboratory studies and seven field studies that passed the minimum standards for review showed an accuracy index (area under the ROC curve) of .85 and .89, respectively.<sup>6</sup> It led the panel to conclude that specific-incident polygraph tests can discriminate lying from truth telling at rates well above chance, though well below perfection (National Research Council, 2003).<sup>7</sup>

The figures mentioned above highlight that the error rate of the CQT can be substantial. Moreover, with the cut-off points used in practice, the test is especially prone to false positive outcomes. This is problematic for application in the legal arena, as it is alien to legal doctrine abbreviated in the so-called Blackstone Maxim: “Better that ten guilty persons escape than that one innocent suffer (Blackstone, 1882; see also Volokh, 1997).”

### *1.3 Legal systems*

The application of the polygraph in different European countries varies considerably. It ranges from a significant role in the judicial system in Belgium to total absence in Spain where the use of polygraph testing has apparently never been suggested. Before we give a concise description of the application of the polygraph in a number of European countries, we should shortly venture into the manner in which courts treat evidence in many European countries.

In most common law countries – like the United Kingdom and most of the states in the United States – the legal system is based on a jury trial. In a jury trial, the judge serves as the gatekeeper with respect to evidence: he decides which pieces of evidence are put before the jury. This has resulted in an extensive body of rules on the admissibility of evidence (e.g., Heydon, 2004; Strong, 1999). In most continental countries however, the criminal

---

<sup>5</sup> These percentages do not necessarily add up to 100% because of the inconclusive category.

<sup>6</sup> The National Research Council did not report accuracy in terms of percentage correct decisions. This is because percentage correct decisions rely on an arbitrary cutoff point. The choice of where the cut-off is placed depends on the preference to reduce either the false positive ratio or the false negative ratio. Rather, the National Research Council expressed accuracy in terms of the area under the Receiver Operating characteristic Curve (ROC *a*).

<sup>7</sup> This conclusion was highly similar to a U.S. government report that was published 20 years earlier (Office of Technology Assessment, 1983). This report concluded that “...the polygraph detects deception at a rate better than chance, but with error rates that could be considered significant.”



## CHAPTER 1

trial is modelled after a bench trial. Here, the judge – or court – is both gatekeeper and finder of fact.

The latter system has resulted in a tradition in which all potential evidence is presented to the fact finder. In this tradition, rules on admissibility of evidence serve little purpose. The emphasis lies not so much on the admissibility of evidence,<sup>8</sup> but on rules governing the use of evidence by the court in decision making. Since these rules tend to be scarce in many countries, this has resulted in a ‘free evidence’ system (Damaška, 1986).

The lack of admissibility rules is usually compensated by the rules that courts – in contrast to juries – have to argue their decisions. Two different versions of this exist in Europe. In the one system, the final evaluation of evidence is governed by the *conviction intime*: the court convicts when it is convinced the suspect is guilty. In the other type – with *conviction raisonnée* – the court must argue its verdict in a manner that conforms to certain statutory standards, although these standards tend to be rather lenient. In all, many European legal systems lack admissibility rules, resulting in a ‘free evidence’ system. As a consequence, policies that forbid the use of polygraph tests as evidence are often of little value, and tests may be presented to the court nevertheless.

### 1.4 The polygraph in Europe<sup>9</sup>

In Belgium, over 300 CQTs are performed annually. Cases include murder, rape, theft, and arson. The introduction of polygraph testing in Belgium raised some legal concern. When questioned in Parliament, the minister of Justice indicated that the results should only be used in police investigations, and would not be offered as evidence in court (Bockstaele, 2000; Verhaegen, 2000).<sup>10</sup> Practice turned out to be different. Although refused as evidence by bench courts (Traest, 2001), results of polygraph tests have also been offered as evidence before the Court of Assisen; the Belgian jury. Since

---

<sup>8</sup> Exceptions are rules on the admissibility of evidence obtained by the police unlawfully.

<sup>9</sup> Our gratitude goes to Anna Baldry, Vivi Bang Pallesen, Wolfgang Bilsky, Vicente Garrido Genoves, Pär Anders Granhag, Helinä Häkkänen, Andreas Kapardis, Martin Killias, Annika Melinder, Pekka Santtila, Sven Svebak, Peter Tak and Rita Zukauskienė, who served as informants for this section.

<sup>10</sup> The minister of Justice wrote: “but it should be emphasized that not so much the results of the test with the lie detector would enable the court to solve a criminal or other case, but the confession to which the test may lead, as far as the confession is believable and supported with other data in the investigation. The lie detector thus is a tool among others; as it has proved useful in certain cases – sometimes incriminating, more often exculpatory – investigating judges will continue to use it in the future.” Our translation from questions and answers Kamer [2000-2001], 22nd May 2001, 12 (question number 4689 by Schoofs).

juries do not have to explain their use of evidence, polygraph tests are *de facto* accepted as evidence in Belgium.

In the Netherlands, the outcome of a polygraph test is neither used as evidence in court, nor in police investigations. In the Nineties, it appeared to go the other way. The Taskforce Lie Detection (Werkgroep Leugendetectie, 1993), consisting primarily of police officials and policy advisors, concluded that the polygraph could have a certain value for the Dutch criminal law system. According to the taskforce, lie detection could give an indication of the veracity of a statement and, as such, could help shape the direction and priorities in criminal investigations. The taskforce, however, did acknowledge the technique's error rates and expressed reservations about using the outcome of a CQT as evidence in court. In 1996, this report was succeeded by two reports on the request of the Dutch minister of Justice by four professors of psychology (Boelhouver, Merckelbach, van Koppen, & Verbaten, 1996; van Koppen, Boelhouver, Merckelbach, & Verbaten, 1996). Their recommendations were generally similar to those made by the taskforce. They concluded that, provided certain conditions are met, application of the polygraph in criminal investigation could be meaningful. Still, the minister of Justice rejected the introduction of polygraph testing, primarily because it is alien to the right to remain silent, an argument that returns in other countries across Europe as well. More recently, a defendant's request for a polygraph test was rejected by the The Hague Appellate Court. This decision was upheld by the Dutch Supreme Court (Hoge Raad) because "it is widely known that the use of polygraphs in criminal investigation is disputed because of its unreliability."<sup>11</sup> In contrast to criminal law, polygraph testing has been introduced in The Netherlands in sex offender management, an application that will be discussed in more detail in chapter 2.

In the United Kingdom, polygraph testing is not used in criminal law either. The British government did announce the intention to initiate pilot studies on its effectiveness in the Nineties, but after a report by the Working Group of the British Psychological Society (British Psychological Society, 1996), the plans were abandoned. More recently, the British Psychological Society published a new report (BPS Working Party, 2004),<sup>12</sup> concluding that "the use of the polygraph has inherent weaknesses, and that

---

<sup>11</sup> See Supreme Court (Hoge Raad der Nederlanden), 18<sup>th</sup> June 2004, *LJN* AU 5496 to be found on [www.rechtspraak.nl](http://www.rechtspraak.nl). See also the The Hague Appellate Court (Hof Den Haag), 14<sup>th</sup> March 2006, *LJN* AP 2846.

<sup>12</sup> The Working Party consisted of Ray Bull, Helen Baron, Gisli Gudjonsson, Sarah Hampson, Gina Rippon and Aldert Vrij.



## CHAPTER I

the error rates can be high. [...] Polygraphic deception detection procedure should not be ascribed a special status” (p. 30). Meanwhile, the United Kingdom is serving as a pioneer in the use of the polygraph in sex offender management.

In Germany, the Supreme Court (Bundesgerichtshof) abandoned the CQT from penal procedures in 1998. This decision was based on a report written by three professors of psychology. There was, however, less agreement among these experts than in the aforementioned countries. According to professor Klaus Fiedler (Fiedler, 1999), the CQT does not meet the standard criteria of scientific validity. Professor Max Steller (Steller & Dahle, 1999) was a little more lenient, acknowledging that polygraph testing should not be completely excluded from forensic application. Psychology professor and practicing polygrapher Udo Undeutsch was most lenient (Undeutsch & Klein, 1999). Nonetheless, based on these reviews, the German Supreme Court forbid CQT outcomes from penal procedures (Burgsmüller, 2000; Fiedler, 1999; Fiedler, Schmid, & Stahl, 2002; Offe, 2001; Steller & Dahle, 1999; Undeutsch & Klein, 1999).<sup>13</sup> Still, the most widespread use of the polygraph in Germany used to be in civil cases, mainly child custody disputes with allegations of sexual abuse of the children. In a recent ruling, however, the Bundesgerichtshof put the outcomes of a polygraph test in these cases on par with those in criminal cases.<sup>14</sup>

In the Scandinavian countries, only Finland has used the polygraph on a relatively large scale. The Finnish National Bureau of Investigation (NBI) has used it in around 300 cases since 1995, mostly homicides and sexual crimes. The NBI is the only institution within the Finnish police using the polygraph, primarily to indicate the direction of investigations. The use of the polygraph at the NBI is managed by two detectives. Even though the outcome of a test is not supposed to serve as evidence in court, evidence in Finland is also governed by the ‘free evidence’ system we discussed above. This has resulted in occasional cases where the outcome of a polygraph test has been presented as evidence to the court. At the moment of writing, one case involving polygraph evidence is under review by the Finish Supreme Court (Korkein Oikeus). This case will likely set a precedent for future cases.

Norway and Sweden are only modest users of the polygraph. In Norway, it has been used in the pre-trial phase in at least three cases during the last 10 years. The admission to use polygraphs as an evidence gathering

---

<sup>13</sup> Bundesgerichtshof 10th February 1999 (3 StR 460/98; *NStZ-RR* 2000, 35).

<sup>14</sup> See Bundesgerichtshof, 24th June 2003 (VI ZR 327/02).

method for the trial itself was questioned, however, and the Norwegian Supreme Court (Høyesterett) rejected the use of polygraphs in 1996.<sup>15</sup> More recent cases have been met with the same scepticism and reference to this verdict. In one of the three cases, the defence made considerable efforts to convince the Appellate Court (Lagmannsrett) to accept polygraph testing as evidence. They failed, and the Court decided on traditional evidence and witness testimony only. Still, the Norwegian evidence rules are also characterized by the 'free evidence' system. As such, the precedent set by the Høyesterett does not prevent use in court because of evidential reasons, but because it violates the civil rights of the suspect. It could, the court argued, be an imperative for others (e.g., other suspects, victims etc.) as well and as such threatens the civil rights and reflect an undue pressure. Apart from criminal cases, the polygraph is also used in civil cases in Norway.<sup>16</sup>

In Sweden, the polygraph is not used on a regular basis either. It was used two or three times during the Nineties in cases where men were accused of molesting children. The results were offered as evidence by the defence, to prove that the suspects were innocent. In one of these cases, the court held a small hearing with one expert pro and one expert con the polygraph. The court decided to give very little weight to polygraph evidence, which had a precedential effect. Therefore, even though there is no formal ban and free examination of evidence, the use of the polygraph is limited. In Denmark, the polygraph is not used at all. It has, however, also not been forbidden on statutory grounds or through precedents set in legal decisions of courts.

In Switzerland, finally, the use of the polygraph is considered an unlawful means of investigation. In this, the Swiss follow Article 6 of the European Convention on Human Rights. The courts interpret this in a manner that nobody can be forced – directly or indirectly – to accuse themselves, even if the suspect complies (Piquerez, 1987; Schmid, 1993).<sup>17</sup>

In sum, the scientific merits of the CQT are heavily debated. Most prominent criticisms levelled against it are the flawed underlying rationale, the lack of theoretical underpinning and the high probability for false positive outcomes. In general, courts around Europe seem to be aware of the controversial status of the CQT. Although many courts have rejected

---

<sup>15</sup> Høyesterett (Supreme Court). Kjennelse (verdict). Rt-1996-1114 (343-96).

<sup>16</sup> Sven Svebak, a university professor, reports that he, to the best of his knowledge, is the only one conducting polygraph examinations in Norway. He conducted some 100 polygraph tests, almost exclusively in civil cases.

<sup>17</sup> Schweizerisches Bundesgericht, 26th April 1994, BGE 118 Ia, that can be found on <http://www.oefre.unibe.ch/law/dfr/a1120031.html>.

## CHAPTER 1

the results of polygraph tests as evidence, due to the 'free evidence' system that governs evidence in most countries, this has almost never been a general ban. Moreover, besides serving as evidence, polygraph test outcomes are also used to direct police investigations, without ever reaching the courtroom



## Chapter 2: Sex offender management using the polygraph: A critical review

Meijer, E.H., Verschuere, B., Merckelbach, H.L.G.J., & Crombez, G. (in press). Sex offender management using the polygraph: A critical review. *International Journal of Law and Psychiatry*.

### **Abstract**

Reducing recidivism is a central goal in the treatment of sex offenders. In Europe, there is an increased interest in using the polygraph (“lie detector”) as a tool in the treatment and risk assessment of convicted sex offenders. This interest originated from optimistic reports by American clinicians who argued that polygraph testing in the treatment of sex offenders is akin to urine analysis in the treatment of drug addiction. In this chapter, we critically examine the validity and utility of post-conviction sex offender polygraph testing. Our review shows that the available evidence for the claims about the clinical potential of polygraph tests is weak, if not absent. We conclude that portraying post-conviction polygraph testing as analogous to urine analysis is inaccurate, misleading, and ultimately, risky.



## CHAPTER 2

### *2.1 Introduction*

Recidivism among sex offenders poses a serious problem. Data from the United States Bureau of Justice Statistics (2003), show that 5% of about 10,000 sex offenders who were released from prison in 1994 were rearrested for a sex crime within three years. At 43%, rearrests for any type of crime were substantially higher. A review of 61 studies ( $n = 23,393$ ) demonstrated that sexual recidivism was 13% after a mean follow-up of 4-5 years (Hanson & Bussiere, 1998). Recidivism for any reoffense was 36%. Moreover, these statistics are an underestimation of true recidivism rates. Follow-up periods during which recidivisms are monitored are by definition restricted, and not all crimes lead to rearrest or reconvictions (Kilpatrick, Edmunds, & Seymour, 1992). Unfortunately, highly effective treatment programs do not yet exist. Meta-analyses on the efficacy of sex offender treatment programs have shown that cognitive-behavioral therapy can decrease recidivism, but that this decrease is only modest (Hall, 1995; Hanson, Gordon, Harris et al., 2002). More recent and more carefully designed studies failed to find any treatment effects at all (Marques, Wiederanders, Day, Nelson, & van Ommeren, 2005; Schweitzer & Dwyer, 2003).

Sex offenders often deny the seriousness and impact of their crimes. Embracing a more truthful and complete report of their crimes is considered an important aspect of treatment. Risk assessment would also benefit from offenders' truthful self-report about their criminal history and pathology. Contemporary risk assessment is based on actuarial risk predictors, which are risk factors derived from empirical research (Hanson, 2003). Hanson and Bussiere's meta-analysis showed that sexual offense recidivism was best predicted by measures of sexual deviancy (e.g., deviant sexual preferences, prior sexual offenses) and, to a lesser extent, by general criminological factors (e.g., age, total prior offenses). Although actuarial risk assessment has improved the prediction of recidivism compared to clinical judgment (Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Quinsey, Harris, Rice, & Cormier, 2006), this prediction remains far from perfect (Hanson, 2003; Harris, Rice, Quinsey et al., 2003). Langton et al. (2007) assessed the predictive ability of six contemporary actuarial instruments in 468 sex offenders who were followed for an average of 5.9 years. Predictive accuracy was expressed in terms of the area under the Receiver Operating Curve (ROC a). This statistic can be interpreted as the probability that a randomly chosen sex offender who reoffended obtained a higher score than a sex offender who did not. Values range from 0 to 1, with 0.5 reflecting chance level. The

## SEX OFFENDER MANAGEMENT USING THE POLYGRAPH

results showed that for all instruments, accuracy was significantly above chance. With ROC a values ranging from .57 to .72 for any crime, and from .61 to .71 for sex crimes, accuracy was, however, only modest.

Offenders' records are used as input for actuarial risk assessment instruments, and, although objective, record information is often incomplete. As an example, the number of prior sex offenses present in the records is probably an underestimation of the true number of sex offenses. Relying exclusively on file information, thus, poses a serious problem, and good practice requires corroborating offenders' self-reports. This, in turn, may prove problematic as offenders have a vested interest in concealing their true behavior, as good behavior is often rewarded with such benefits as parole. In this context, some authors have referred to offenders' fake good tendencies as "supernormality" (Cima, Merckelbach, Hollnack et al., 2003). This supernormality is a major obstacle for accurate risk assessment. All together, both treatment and risk assessment in sex offenders could benefit from more honest self-reports.

In this review, we critically evaluate a method for verifying the veracity of offenders' self-reports; post-conviction sex offender polygraph testing (PCSOT). This method has steadily gained popularity among clinicians over the last decade, and was used in 70% of community sexual abuser programs for adult males in 2002 in the United States (McGrath, Cumming, & Burchard, 2003). Proponents argue that this method contributes to a more accurate and more complete sexual history taking, more accurate offense descriptions, superior assessment of high risk behaviors, and better compliance with probation conditions. The popularity of the method among practitioners is illustrated in the opinion expressed by the former chair of the American Polygraph Association: "Those of us who conduct post-conviction sex offender testing know that it has been the missing link in preventing recidivism (Consigli, 2002, p. 239)". American practitioners have presented the role of the polygraph in the treatment of sex offenders as similar to that of urine analysis in the treatment of drug addicts (English, Jones, Patrick, & Pasini-Hill, 2003). Both are viewed as objective and accurate methods to verify whether the offender's reports are truthful. These positive claims received much attention, and against this background, PCSOT was recently introduced in Europe. In the United Kingdom and The Netherlands, maximum secure treatment hospitals have started PCSOT studies. European users appear to share Consigli's enthusiasm, promoting widespread use. At the time of writing, in the United Kingdom, the Management of Offenders and Sentencing Bill was in progress, extending the use of polygraph tests for sex offenders on license.



## CHAPTER 2

In this review, we discuss whether this enthusiasm is justified. We will first describe the general principles of polygraph testing and how it is used in PCSOT. We will then critically examine the results from PCSOT studies. Finally, we will discuss whether there is any evidence for the notion that PCSOT reduces recidivism, and whether PCSOT is, indeed, the ‘urine test’ of sex offender treatment.

### *2.2 The polygraph and the detection of deception*

The type of polygraph test used in PCSOT is a version of the Comparison Question Technique (Grubin, Madsen, Parsons, Sosnowski, & Warberg, 2004). This type of test is also known as the Control Question Test (CQT; National Research Council, 2003; Reid, 1947), and is best known for its use by law enforcement agencies in criminal investigations. During a CQT, relevant and comparison questions are asked. The relevant questions are directly related to the incident under investigation (e.g., “On the 14<sup>th</sup> of May, did you kill X?”). The comparison questions are emotionally provocative, and deal with undesirable behavior that is unrelated to the crime itself (e.g., “During the last 25 years, have you ever done anything illegal?”). Physiological responses, including skin conductance response, respiration, and blood pressure are monitored throughout. The test procedure is based on the assumption that the questions that pose the biggest threat to the individual will elicit the strongest physiological responses. It is further assumed that the relevant questions are most threatening for guilty suspects, whereas comparison questions form the biggest threat to innocent suspects. This latter assumption requires some explanation. A CQT is preceded by a lengthy interview, in which the examinee is led to believe that the polygraph is almost infallible. Thus, the innocent examinee can confidently and honestly answer “no” to the relevant questions. Furthermore, the examinee is led to believe that a deceptive answer to any of the questions will lead to a ‘deceptive’ test outcome. At the same time, the examinee is pressured into answering “no” to the comparison questions by the examiner suggesting that confessing illegal activities will negatively influence the examiner’s opinion (see e.g., Offe & Offe, 2007). As a result, the innocent examinee is assumed to show the strongest physiological responses to the comparison questions, fearing that his/her deceptive answer to this question will get him/her convicted for the crime under investigation.

### *2.3 Polygraph validity*

Ever since its introduction, the CQT has been the object of an intense debate, and remains controversial (Ben Shakhar, 2002; Ben Shakhar & Furedy, 1990; Fiedler et al., 2002; Furedy, 1996a, 1996b; Honts, Kircher, & Raskin, 1995; Lykken, 1998; Raskin & Honts, 2002). The debate focuses on the assumptions underlying the CQT, thus calling into question the test's accuracy. Many critics have argued that it is impossible to determine the origins of the physiological responses. Stronger responses to the relevant questions may indicate a guilty suspect's fear of being detected, but may also indicate an innocent suspect's fear of a wrongful conviction (Lykken, 1998). Indeed, whether an innocent examinee will be worried most by the comparison questions heavily depends on the skills of the polygrapher. Additionally, the precise detection accuracy of the CQT is hard to establish. In laboratory studies, participants are not threatened by severe consequences of failing a test. The studies therefore lack ecological validity. Field studies, on the other hand, suffer from a lack of "ground truth", i.e., conclusive exonerating or incriminating evidence that can corroborate test outcome (Iacono, 1991).

A large number of studies addressed the accuracy rates of the CQT in specific incident testing (e.g., tests in criminal investigations). Mock crime study estimates range from 74-80% for guilty participants, and 63-66% for innocent participants (Ben Shakhar & Furedy, 1990; Kircher et al., 1988). The false positive ratio (i.e., innocent examinees classified as guilty) was estimated at 12-15%, and the false negative ratio (i.e., guilty examinees classified as innocent) was 7-8%. Field study estimates of the accuracy of the CQT are slightly higher, ranging from 84-89% for guilty, and 59-72% for innocent suspects, with estimated false positive rates of 12-23% and false negative rates of 1-13% (Ben Shakhar & Furedy, 1990; Raskin & Honts, 2002).

Recently, the National Research Council (NRC; 2003) evaluated the validity of the CQT for screening purposes. These tests are used in the United States to screen job applicants and monitor employees of government agencies whose work involves security risks (e.g., FBI applicants or nuclear scientists; Krapohl, 2002). When reviewing the evidence, the NRC concluded that there were no high quality studies on the use of the CQT for screening purposes. Therefore, the council based its review on studies that addressed the validity of the CQT in specific-incident testing. Reviewing this evidence, the NRC noted that most studies do not reach high levels of scientific quality. For example, nearly all studies are based on naïve participants, threatening ecological validity. Still, the Council



## CHAPTER 2

selected 37 laboratory studies and seven real-life field studies that met the minimum criteria for its analysis. The Council reported a median ROC *a* of .86 for the CQT in specific-incident testing, and concluded that specific incident polygraph tests can discriminate lying from truth telling at rates well above chance, though well below perfection. Furthermore, the Council pointed out that accuracy for screening purposes is almost certainly lower than what can be achieved for specific-incident testing. The latter is especially relevant because the use of the CQT in PCSOT predominantly<sup>18</sup> involves questioning sex offenders about unknown incidents, thus bearing a stronger resemblance to screening than to specific incidence testing (Abrams & Abrams, 1993). The main difference between specific incidence and screening tests is that during a specific incident polygraph test, a suspect is questioned about a *known* incident, such as a murder, theft or arson. In a screening test, however, someone is questioned about incidents of which it is *unknown* whether they have taken place, such as disclosing classified information.

### *2.4 The polygraph and sex offenders*

An important reason for lower accuracy in screening applications than in specific incident testing is that the former involves considerably more ambiguity for the examinee (National Research Council, 2003). This becomes clear when looking at the function of the CQT in PCSOT more closely. PCSOT intends to verify self-reported information about past or current behavior. Firstly, it entails verifying the accuracy and completeness of the sexual history information provided by the offender. This “sexual history disclosure test” is generally administered at the beginning of the treatment. Secondly, it entails verifying whether a probationer or parolee is complying with the conditions of community supervision and is cooperating with treatment expectations (English et al., 2003). Such “maintenance tests” are administered periodically, for example every six months (Abrams & Abrams, 1993). Thus, in contrast to specific-incident testing, the offender is questioned about incidents of which it is unknown whether they have taken place at all. This means that the relevant questions need to be phrased in a very broad way. Consequently, relevant questions (e.g., “Have you had unsupervised contact with children over the last 3

---

<sup>18</sup> There are some exceptions to this rule: Some authors (e.g., Abrams & Abrams, 1993; English et al., 2003) also report the use of specific-incident testing in the treatment of sex offenders, for example, when the offender continues to deny committing the crime of conviction.

## SEX OFFENDER MANAGEMENT USING THE POLYGRAPH

months?") and comparison questions (e.g., "Have you done anything over the last 3 months that would concern your probation officer?"; Grubin et al., 2004, p. 213) become more similar. As a result, test outcomes will be more ambiguous and diagnostic decisions less accurate (see also Ben-Shakhar, in press).

We found no studies that made a serious attempt to estimate the accuracy of the CQT in PCSOT. The only studies that tried to estimate it limited themselves to offenders' self-reports of test accuracy. Kokish, Levenson, and Blasingame (2003), for example, collected self-report data from 95 sex offenders who underwent a total of 333 tests. Eighteen offenders (19%) reported having been incorrectly labeled deceptive, and 6 (6%) claimed they had incorrectly been found truthful. Similar statistics were reported by Grubin and Madsen (2006). They collected self-report data from 126 offenders who completed a total of 263 tests. Of these 126 offenders, 27 (21%) reported having been incorrectly found deceptive, and 6 (5%) reported they had been wrongly labeled truthful.

Although such results might look encouraging, these studies suffer from a major methodological pitfall, namely sampling bias. Participation in a polygraph treatment program is voluntary. In a prospective study by Grubin and colleagues (2004), only 21 out of the 116 offenders (18%) that were initially approached, completed both polygraph tests that were part of the study. Accuracy rates based upon this small sample of offenders may thus very well be an overestimation. Offenders that were confronted with an incorrect test result may have simply dropped out.

Clearly, more thorough validity research on the CQT in PCSOT is needed. Apart from the ambiguity of the relevant questions, there are a number of other factors that specifically threaten PCSOT validity (see also Branaman & Gallagher, 2005; Cross & Saxe, 1992, 2001; Faller, 1997). These factors are related to cognitive distortions of sex offenders, and the effects of repeated testing.

Pertaining to cognitive distortions, sex offenders are known to justify their sexual involvement with children in diverse ways. They may minimize harm caused by the abuse, perceive children as desiring sexual contact with adults, or perceive their sexual contact as socially acceptable (Gannon & Polaschek, 2006; Ward, Hudson, Johnston, & Marshall, 1997). In specific-incident testing, the effects of cognitive distortions can be minimized by asking very specific and straightforward questions. For example, the suspect might be asked "On the night of July first 1994, did you put your penis into Sarah's vagina?" As argued earlier, in PCSOT one cannot ask questions as specific as this because the examiner does not know whether deviant



## CHAPTER 2

behavior actually took place. The relevant questions are therefore formulated more vaguely; e.g., “Since July first 1994, have you had any sexual contact with a minor?” (Hagler, 1995, p. 107). This type of question is problematic when dealing with sex offenders, because sexual contact is not explicitly defined. It allows interpretation, and in the distorted mind of the sex offender, “sexual contact” per se may not have taken place. It may, for example, have been labeled as normal father to daughter interaction. A sex offender who has cognitive distortions of this type may therefore not respond to relevant questions. It is easy to see how this may contribute to false negative test outcomes.

The second factor that threatens CQT accuracy in PCSOT is repeated testing of the same offender. This can deteriorate the test outcome for a number of reasons. To begin with, it is well known that physiological responsivity decreases upon repeated presentation of the same stimulus (i.e., habituation; Andreassi, 2000). Thus, the ‘guilty’ sex offender may not show a marked physiological response to the relevant question after having been repeatedly confronted with it, thereby raising the probability of a false negative test outcome. Repeated testing is also likely to reduce the emotionally provocative loading of the comparison question, thereby increasing the probability of a false positive test result in ‘innocent’ sex offenders. Furthermore, repeated testing increases the probability of the effective use of countermeasures by the offender. Countermeasures refer to everything the examinee can do in order to alter the test outcome (Honts & Amato, 2002). The examinee might use physical countermeasures such as biting of the tongue in order to create a physical response to the comparison question. Alternatively, he or she might try to alter physiological responses using mental countermeasures, by thinking about something exciting during the comparison question, for example. Several studies have shown that mental countermeasures are especially problematic because they are not easily detected by the examiner (for a review, see Honts & Amato, 2002). Honts (2004) has suggested that countermeasures may only be effective if the examinee is able to practice them. Repeated testing not only familiarizes the examinee with the procedure, it also gives the examinee the opportunity to practice countermeasures.

In sum, the validity of the CQT has been, and still is, debated. The accuracy of this test for specific incident testing is far from perfect, and when used in PCSOT, matters worsen even more. High-quality research on the use of the CQT for screening purposes is non-existent, and there is ample reason to assume that its accuracy will not reach the level obtained in specific-incident testing. Furthermore, several factors (i.e., cognitive

## SEX OFFENDER MANAGEMENT USING THE POLYGRAPH

distortions, repeated testing and countermeasures) seriously threaten its validity in PCSOT.

### *2.5 Disclosure through polygraph testing*

Some have argued that the polygraph has its utility, regardless of its accuracy. "I am less concerned about the reliability of polygraph. The fact is that every time I refer a client for polygraph testing, I end up with significantly more information than I started with" (Veeder, 1995, in Hagler, 1995). Several reports have appeared, suggesting that the polygraph is highly successful in obtaining previously undisclosed information (e.g., Ahlmeyer, Heil, McKee, & English, 2000; Emerick & Dutton, 1993; English et al., 2003). This includes information on the number of previous offences, age at the first offence, number of victims, and prevalence of high risk behaviors.

In this particular respect, recent European PCSOT pilot studies also report encouraging results. Among them, Wilcox and Sosnowski (2005) describe the results of 14 sex offenders on probation who were administered a single CQT. Important new information was obtained from all offenders: 93% of the sample admitted they had committed both contact (e.g., indecent assault) and non-contact (e.g., voyeurism) sexual offenses, compared to the known 29% in the probation records. In addition, the offenders reported earlier onset of offending and a wider range of paraphilic interests than had previously been reported. The prospective study by Grubin and his colleagues (2004) also demonstrated that 97% of the offenders disclosed high-risk behavior at their first test, and 71% did at the second test. Preliminary results from a Dutch pilot study also showed an increase in the disclosure of number of offenses, high-risk behavior and paraphilic interest in a sample of 23 men convicted of child pornography possession (Buschman, personal communication, October 28, 2006).

Overall, there is little doubt that PCSOT leads to an increase in the amount of information disclosed by offenders. There are, however, a number of methodological points that qualify the optimistic interpretation that one might give to this effect of polygraph testing. First, as discussed earlier, participation in a treatment program involving PCSOT is voluntary. Thus, the offenders included in the studies cited above probably represent a subgroup of highly motivated (and perhaps quickly disclosing) offenders. Second, the ground truth is unknown. Thus, it remains unknown how much information has gone undetected and how accurate the newly disclosed information is. As acknowledged by several authors (Grubin & Madsen, 2006; Kokish, Levenson, & Blasingame, 2005), offenders might have fabricated stories after deceptive test outcomes, in order to satisfy



## CHAPTER 2

examiners or to obscure the actual reason for failing the test. Finally, all studies so far lack a proper control condition, making it impossible to draw firm conclusions on the utility of the polygraph per se. The additional value of PCSOT in obtaining new information can only be established by comparing its effect to that of a control condition such as simply interviewing offenders. Researchers have now compared information obtained through PCSOT solely with what was already known from the offender's files. In this context, any newly disclosed information is reported as evidence for the success of PCSOT. Yet, this approach remains silent about whether a thorough interview might have been sufficient to obtain this information and whether the polygraph is additionally needed. Thus, the superiority of PCSOT can only be demonstrated by comparing it to another method that has a similar intensity (i.e., similar amount of time spent with the examinee). Only such a comparison could demonstrate the additional value of PCSOT in obtaining new information. If future research were to demonstrate that this is the case, then the question arises as to where this effect comes from. There is ample reason to assume that it is related to the bogus pipeline effect.

### *2.6 Bogus pipeline*

There is a long tradition in social psychology to develop methods that aim to reveal a person's true opinion on sensitive topics. Standard paper-and-pencil tests are believed to be unreliable, because they are subject to response biases such as social desirability (Crowne & Marlowe, 1964). One approach social psychologists have used to circumvent social desirability is attaching the person to an impressive but useless machine. People respond more truthfully when they are led to believe that this "bogus pipeline" can reveal the actual truth. Hooked up to a fake lie detector, people have been found to be more honest about social prejudices, sexual fantasies, and addictive behaviors (Roese & Jamieson, 1993). Likewise, the capability of the polygraph to obtain new confessions in sex offenders is probably more related to this bogus pipeline effect than to its accuracy. Some proponents of sexual history disclosure tests have acknowledged this. According to Abrams and Abrams (1993), there are three points in time when sex offenders can disclose information: (1) when they are told that they will face a test in the near future, (2) during the pre-test interview, and (3) during the confrontation with the test outcome. Note that the first two points are before the polygraph test is actually conducted. Indeed, Grubin and his colleagues (2004) found that most information was disclosed during the pre-test interview, well before the sensors of the polygraph were actually

## SEX OFFENDER MANAGEMENT USING THE POLYGRAPH

connected. Apparently, the expectation of an upcoming polygraph test is sufficient to make offenders disclose information. This means that disclosure of new information may have little to do with the diagnostic accuracy of the polygraph. More likely, it is the questioning and the intimidation by the lie detector that makes the offender bring up new information. Recent empirical data supports the bogus pipeline effect in sex offenders. Gannon, Keown, and Polaschek (2007) asked 41 child molesters to what extent they agreed with cognitive distortions such as "Sometimes, touching a child sexually is a way to show love and affection" (MOLEST; Bumby, 1996). All items were read out loud and the participants gave a verbal response. Four to six weeks later, this procedure was repeated, but now 18 participants were attached to a bogus lie detector when answering the items. The remaining 23 participants formed a control condition in which the procedure was identical to the first administration. Results showed that child molesters in the bogus pipeline condition, but not in the control condition, reported more cognitive distortions than in the first administration, confirming the success of lie detectors in obtaining more honest self-reports, regardless of their accuracy. Whether the polygraph should be used to extract confessions regardless of demonstrated accuracy is an ethical dilemma.

Another ethical dilemma PCSOT faces is that sex offenders are required to be completely honest for successful completion of their treatment program. This requirement stands in sharp contrast to the position of the clinician who, when using the CQT is relying on a test that is based on deception. That is, the examinees are erroneously told that a deceptive answer to one of the comparison questions will lead to a deceptive test outcome (Offe & Offe, 2007). When discovered, this manipulative behavior of the clinician may seriously affect the therapeutic relationship.<sup>19</sup>

---

<sup>19</sup> Besides the ethical issues, PCSOT also faces legal issues. The question arises on how to handle criminally-relevant information that is disclosed by the offender. There are two options: either the offender is prosecuted or he/she gets immunity. When prosecuting the offender for the newly disclosed information, it is likely that defense attorneys will challenge the charge because it is based on self-incriminatory evidence that violates the right to remain silent and/or the right to have counsel present during interrogative questioning (see Malesky & Packer, 2004). Likely, prosecuting will imply the end of the treatment in most cases. Moreover, it is likely that offenders who risk being charged will not participate in polygraph-supported treatment programs. The other possibility is to promise the sex offender immunity, in that no legal consequences will be attached to the newly disclosed criminal information. This is the strategy that is most often chosen (Schwartz & Cellini, 1995). However, this strategy may prove very difficult to explain to the victims of the disclosed offences.



## CHAPTER 2

### *2.7 Reducing recidivism*

The most critical question is not to what extent PCSOT leads to the disclosure of information, but whether it reduces recidivism. Such reduction could be either direct, due to the deterrent effect of an upcoming polygraph on behavior, or indirect through increased disclosure improving risk assessment and/or treatment outcome. We will argue that there is evidence against the former, and no evidence for the latter.

PCSOT has been used since the 1960s, and has been described as the missing link in preventing recidivism. Going through all relevant scientific databases (i.e., Psychinfo, Medline, Web of Science), however, we could not identify a single study that directly addresses the critical issue of whether PCSOT actually reduces recidivism.<sup>20</sup> There is little reason to assume that PCSOT will have a direct effect on recidivism. Consider, for example, the study by Grubin and colleagues. These authors investigated whether the prospect of a polygraph test would influence offenders' behaviors. Clearly, the authors expected that the prospects of a polygraph test would result in a reduction in high risk behavior (Grubin et al., 2004). One-hundred and sixteen convicted sex offenders were approached, of which less than half ( $n = 50$ ) agreed to participate. High risk behaviors were identified for each individual and two groups were created. One group was told that they would undergo a polygraph test in three months, whereas the other group was told their behavior would be reviewed, with no specific reference to the use of the polygraph. In fact, both groups were subjected to a polygraph test three months later. Thirty-two of the original 50 offenders reported for the polygraph test. Thirty-one of them (97%) disclosed having engaged in high risk behavior. Importantly, there was no difference between the polygraph aware and the polygraph unaware group. For a second polygraph test, performed three months later, another 11 offenders dropped out. Of the remaining 21 offenders, 15 (71%) disclosed high risk behavior. This study shows that although PCSOT is able to reveal new information in a highly selected sample of motivated sex offenders (18% of the original sample completed both tests), the knowledge of an upcoming polygraph test does not seem to prevent offenders from engaging in high risk behavior.

---

<sup>20</sup> Recently, a study was published that found no difference in recidivism rates between a group with, and a group without PCSOT. This study is discussed in more detail in the general discussion.

## ***2.8 Concluding remarks***

Treatment of sex offenders is a serious issue. Not only does recidivism occur frequently, the public in many countries is highly sensitive to the topic (Lösel, 2000). With this in mind, it is understandable that practitioners resort to unorthodox techniques such as polygraph testing. At present, however, there is no evidence supporting the accuracy of the CQT in PCSOT. Relevant theory and research suggest that this accuracy will be lower than that obtained in criminal investigations. Users have argued that its actual accuracy is less important than the utility of the polygraph in obtaining disclosures. We have argued that this effect has less to do with the actual outcome of the polygraph test, but can be accounted for by its intimidating effect. Moreover, the limited evidence available suggests that this effect does not deter sex offenders from engaging in high risk behavior.

Proponents of PCSOT have argued that it is akin to urine analysis in the treatment of drug addiction. It should be clear from our review that this analogy is seriously flawed. Unlike urine analysis, PCSOT involves deception of the examinee. PCSOT is not as standardized as urine analysis, and test scoring of PCSOT protocols requires far more interpretation. Unlike urine analysis, accuracy of the CQT in PCSOT is unknown and it is highly likely that errors will frequently occur. If false positives occur, the false accusations that follow from these may be highly damaging to the therapeutic relationship. Even more problematic are false negatives: some sex offenders at risk for recidivism will be given a wild card by the polygraph.

PCSOT advocates seem to have unrealistic confidence in the polygraph. This confidence creates the possibility for offenders to engage in sex crimes under the safe conduct of a “non-deceptive” polygraph result. Therefore, we argue against the use of sanctions and privileges for deceptive and non-deceptive polygraph results, as proposed by Ahlmeyer et al. (2000). The occurrence of errors is especially problematic because, unlike urine drug analysis, accurate retesting is impossible with PCSOT. We seriously doubt that clinicians would use urine analysis in the treatment of drug addiction if it would have the same characteristics as PCSOT.



# Part II

## **Chapter 3: Concealed information**



## CHAPTER 3

### *3.1 Introduction*

The first part of this dissertation concerned the widely used Control Question polygraph Test. There does exist a second forensic application of the polygraph that is less popular among practitioners. This application was first described by Münsterberg (1908) and later by David Lykken (1959) who named it the Guilty Knowledge Test. It is nowadays also referred to as the Concealed Information Test (CIT). During a CIT, test questions do not directly address the incident under investigation. Rather, all questions concern details of the crime, presumably only known to the police and to the perpetrator. Answer alternatives to these questions are presented serially, while physiological signals are recorded. These answer alternatives include the correct answer, but also several plausible but incorrect answers (e.g., “Was the victim killed with a ... (a) gun, (b) knife, (c) rope, (d) bat, (e) ice pick?”). For an innocent suspect, all alternatives are equally plausible, and will elicit similar physiological responses. For a guilty suspect, on the other hand, the correct alternative is salient, and will elicit an enhanced physiological response. When multiple questions, each pertaining to different details, are presented to the suspect, and he or she shows a pattern of stronger responding to the correct alternative, knowledge of intimate details of the crime is determined, from which involvement can be inferred.

One of the main advantages of the CIT is that the probability of a false positive test outcome is fully under the control of the examiner. A false positive test outcome means that, merely by chance, a pattern of stronger responding to the correct alternatives has occurred. The probability of this happening depends on two factors. The first factor concerns the test’s properties; the false positive probability is inversely related to the number of questions and the number of answer alternatives per question. The second factor that determines false positive probability is how one defines ‘a pattern of stronger responding’. When a guilty test outcome requires the suspect to respond maximally to the correct alternatives of all five questions, the probability of this happening by chance, is smaller than when one requires a maximal response for only three out of the five correct alternatives. This control over false positive probability has important implications. For one thing, it allows the examiner to set the false positive probability at an arbitrary low level, as prescribed by legal doctrine in most civilized countries.<sup>21</sup> Relatedly, it allows for calculation of the probability

---

<sup>21</sup> This is often abbreviated in the so-called Blackstone Maxim: “Better that ten guilty persons escape than that one innocent suffer.” See Blackstone (1882).

that a guilty test outcome is incorrect. This is essential information if an incriminating test outcome is introduced in court proceedings, as it allows for proper weighing.

### ***3.2 Skin Conductance Response***

The CIT originally described by Lykken (1959) used Skin Conductance Response (SCR) as the dependent measure. This measure has by far received the most attention in CIT research, and has been shown robust in discriminating between guilty and innocent participants. Elaad (1998), for example, reviewed 15 mock crime studies and found average detection rates of 81% for guilty examinees and 96% for the innocent. Similar accuracy rates were reported by Ben-Shakhar and Furedy (1990), who reviewed 10 mock crime studies and found detection rates of 84% and 94% respectively. A more recent review showed similar results with successful detection of 76% of participants with concealed knowledge and slightly lower detection, 83%, of those without (MacLaren, 2001). The two available field studies (Elaad, 1990; Elaad, Ginton, & Jungman, 1992) show equally high detection of innocent suspects (98% and 94%, respectively), but somewhat lower detection accuracy among guilty suspects (42% and 76%, respectively). The latter was, at least in part, attributed to the suboptimal format of the tests, as they contained only a low number of questions. The most extensive meta-analysis so far was performed by Ben-Shakhar and Elaad (2003). Eighty studies were included, yielding an average effect size ( $d$ ) of 1.55. This is almost twice the lower limit of what is typically considered a large effect size (Cohen, 1988). Furthermore, an area under the Receiver Operating Characteristic curve of .82 was reported. This effect size and area under the Receiver Operating Characteristic curve were even higher under conditions that best mimicked real life application (3.12 and .95, respectively). Finally, in its 2003 report, the National Research Council selected 13 studies, yielding an area under the Receiver Operating Characteristic curve of .88.

In all, the available data suggests that at levels of high specificity, the CIT is characterized by acceptable sensitivity. This highlights how the CIT can be improved: by increasing sensitivity while maintaining high specificity. One way this can be achieved is by replacing or expanding the number of parameters that are recorded. Measures other than skin conductance that have been shown successful in a CIT include Respiration Line Length (Ben-Shakhar & Dolev, 1996), pupillary size (Lubow & Fein, 1996), eye blinks (Fukuda, 2001), reaction times (Seymour, Seifert, Shafto, & Mosmann, 2000), finger pulse waveform length (Elaad & Ben-Shakhar, 2006), fMRI (Langleben, Schroeder, Maldjian et al., 2002), imaging of the facial skin



## CHAPTER 3

surface temperature (Pollina, Dollins, Senter et al., 2006), and the P300 component of the ERP (Farwell & Donchin, 1991). Especially the latter has received considerable attention.

### **3.3 P300**

The P300 is a positive brain potential that occurs between 300 and 800 ms after stimulus onset. It was first described by Sutton and his colleagues in the mid 60s (Sutton, Braren, Zubin, & John, 1965), and has appeared in thousands of published experiments over the last 40 years. Typically, P300 based research utilizes some kind of oddball paradigm. In this paradigm, a P300 is elicited by rare target stimuli (e.g., high tone), presented in a series of standard stimuli (e.g., low tones). The participant is instructed to direct attention to the rare target tone, for example by keeping count of all high tones. As the P300 amplitude is dependent on cognitive processing of the stimulus rather than its physical properties, it is often referred to as an endogenous component (Luck, 2006).

It is this endogenous nature of the P300 that makes it exceptionally useful for the detection of concealed information. The underlying rationale is that because the P300 is elicited by rare and meaningful stimuli, only in guilty suspects will the correct alternative elicit a P300 when it is presented among equally plausible, but incorrect alternatives. This approach was first adopted in the late eighties (Farwell & Donchin, 1991; Rosenfeld, Cantwell, Nasman et al., 1988), and since that time, many researchers in the field have emphasized its strengths (Ben Shakhar, Bar Hillel, & Kremnitzer, 2002; Elaad, 1998; Iacono, 2007, 2008). Accordingly, in its 2003 report, the National Research Council recommended further investigation of measurement of event-related potentials as an alternative or supplement to the polygraph (National Research Council, 2003).

Table 3.1 lists the accuracy of P300 based CIT studies. Mock crime studies yielded an average sensitivity (i.e., the proportion of correctly classified guilty participants) of 70%, with a false positive rate of 8%. When autobiographical information served as stimulus material and participants were instructed to deny recognition of self referent-stimuli, for example of their name or birth date, sensitivity reached 87% while false positive rate was 9%. The use of recently memorized stimulus material<sup>22</sup> resulted in an

---

<sup>22</sup> The type of information that is memorized differs between studies. Allen, Iacono, and Danielson (1992) and Ellwanger, Rosenfeld, Sweet, and Bhatt (1996) gave the participants a list with words to memorize, and then instructed them to conceal recognition of the learned words. Ellwanger and colleagues (1996), as well as Rosenfeld, Biroshak, and

## CONCEALED INFORMATION

average sensitivity of 80%, with a false positive rate of 3%. The last group of studies uses information from real life. Here, stimuli refer to events that the participant has, or has not experienced or performed in the past, for example the use of a fake ID (Johnson & Rosenfeld, 1992). Ground truth is established by means of a questionnaire before or after the experiment. The average sensitivity in this group of studies was 80%, with a false positive ratio of 7%. The mean sensitivity over all studies was 80%, with a mean false positive rate of 7%, by which accuracy rates obtained with the P300 are similar to those obtained with SCR.

Thus, criterion validity for the P300 in the detection of concealed information has been established. Besides criterion validity, however, incremental validity is also of importance. This term refers to whether concealed information is detected more accurately by a new measure, than by other measures normally used in the same investigation (see also Sechrest, 1963). Given the amount of research conducted, combined with the relative ease and low costs of measurement, SCR is the most likely benchmark measure for the detection of concealed information. Accordingly, it is relevant whether new measures, either alone or combined with SCR, outperform SCR alone (for examples see Ben-Shakhar & Dolev, 1996; Elaad & Ben-Shakhar, 2006; Gronau, Ben-Shakhar, & Cohen, 2005; Lubow & Fein, 1996). The incremental validity of P300 above and beyond SCR has yet to be established (see chapter 5).

### ***3.4 SVT***

When combining measures, incremental validity is inversely related to the correlation between measures. Highly correlated measures will successfully detect concealed information in the same individuals, but will also miss the presence of concealed information in the same individuals. This makes measurement of both superfluous. An index for the detection of concealed information that is most likely to be uncorrelated to SCR can be derived from symptom validity tests (SVT). SVT refers to a set of neuropsychological techniques aimed at exposing patients who simulate cognitive or perceptual deficits. SVTs generally have a forced choice format, with the underlying rationale that patients with a true impairment will show

---

Furedy (2006) used the experimenter's name as stimulus material. I also included the oddball version of the match to sample task used by Ellwanger, Rosenfeld, Hankin, and Sweet (1999) and Rosenfeld, Ellwanger, Nolan et al. (1999) in this category. In this task, first, a sample number is presented on the screen, after which a sequence of numbers is presented, of which only one matches the sample. Participants are instructed to feign a short term memory deficit, and deny recognition of the matching sample.



## CHAPTER 3

a random response pattern. Any deviation from randomness indicates intact cognition or perception. Pankratz (1983), for example, used the following SVT to detect a feigned memory deficit: the patient was seated in front of two lights, one red and one white. Then, one of these two lights was turned on for 2 seconds, and the patient was asked to remember which one. Fifteen seconds later, the patient was asked to indicate whether the red or the white light had been turned on. This whole procedure was repeated many times. In such a test, patients with a genuine short-term memory deficit will perform at chance level, i.e., approximately half of their answers will be correct. If performance falls significantly below chance level, however, malingering can be inferred. That is, it can be concluded that the patient did remember, but actively avoided the correct answer in order to appear genuinely impaired.

The SVT has also been shown useful in the detection of malingered amnesia for a crime. Denney (1996) described three cases where the defendant claimed amnesia for their crime. In all cases, defendants performed well below chance level on an SVT consisting of questions concerning intimate details of the crime (e.g., “How did the perpetrator leave the bank? (1) walking, (2) running”). More recent research on the accuracy of the SVT in detecting feigned amnesia for mock crimes shows that, at specificity levels of 95 %, its sensitivity ranges from 40% to 60% (Jelicic, Merckelbach, & van Bergen, 2004a, 2004b; Merckelbach, Hauer, & Rassin, 2002). Because the SVT is based on binomial theorem, it shares essential characteristics with the CIT. Most importantly, false positive probability is under control of the experimenter. As is the case with the CIT, it is determined by the number of questions, the number of alternatives, and the cutoff point for guilt. In principle, this would make it a cost effective alternative or supplement to a skin conductance based CIT, but this needs to be explored in empirical studies (See chapter 6).

### ***3.5 Outline***

In the following three chapters, I will present empirical data, all related to improving the detection efficiency of the CIT. Chapter 4 reports two experiments in which we tested the validity of the P300 in the detection of concealed face recognition. In chapter 5, the incremental validity of the P300 beyond that of SCR, and their relation with psychopathic traits is tested. In chapter 6, the SVT is tested as a potential tool for the detection of concealed information.

Table 3.1. Overview of the P300 based CIT studies.

	N guilty	Sensitivity	False Negatives	N innocent	Specificity	False Positives	Notes
<b>Mock crime probes</b>							
Farwell & Donchin (1991) Experiment 1	20	90	0	20	85	0	
Rosenfeld et al. (1988)	7	100	0	5	100	0	
Rosenfeld et al. (2004) Experiment 1	11	82	18	11	91	9	Sensitivity dropped to 18% when participants used countermeasures
Abotalebti et al (2006)	59	58 - 73	27 - 42	51	83 - 85	15 - 17	Hit rates were dependent of scoring method
Mertens & Allen (2008)	16	27 - 47	13 - 53	16	44 - 100	0 - 6	Hit rates were dependent of scoring method. Sensitivity dropped to 7% - 27% when participants used countermeasures.
Rosenfeld et al (2007)	21	43	57				Mock crime condition only, averaged over one probe and multiple probe conditions.
(Weighed) average	134	69.9	23.9	103	88.7	8.4	Average contains scorings methods yielding the highest accuracy.
<b>Autobiographical probes</b>							
Rosenfeld et al. (1995) Experiment 1	13	87	13				Sensitivity for Experiment 1 is the average of the three blocks.
Experiment 2	14	93	7				
Soskins et al. (2001) Experiment 1	13	100	0				Hit rates are based on the p-p measure op P300.
Experiment 2	23	91	9	23	87	13	
Miller et al. (2002)	14	93	7				Hit rates are based on the p-p measure op P300.
Rosenfeld et al. (2004) Experiment 2	13	92	8				Sensitivity dropped to 50% when participants used countermeasures.
Control condition	10	90	10				
Miller & Rosenfeld (2004)	24	96	4				Average over high and low psychopathy scores.
Elhwanger et al. (1996) Experiment 1	26	88	12				p - p measure, averaged over naive and sophisticated malingers. For Experiment 1, only birthday block is given.
Experiment 3	13	69	31	11	100	0	
Rosenfeld et al. (2006)	21	86	14				Subject name condition only, averaged over target and no target condition.
Rosenfeld et al. (2007)	22	59	41				Autobiographical condition only, averaged over one probe and multiple probe conditions.
(Weighed) average	206	86.5	13.5	34	91.2	8.8	



Table 3.1. Continued.

	N guilty	Sensitivity	False Negatives	N innocent	Specificity	False Positives	Notes
Learned probes							
Rosenfeld et al. (1999)	15	87	13	15	100	0	March to Sample task.
Rosenfeld et al. (2006)	21	52	48				Experimenter name condition only. Averaged over target and no target condition.
Allen et al. (1992)							
Experiment 1	20	95	5	20	96	4	
Experiment 2	20	93	7	20	94	6	
Experiment 3	20	95	5	20	98	2	
Ellwanger et al. (1996)							
Experiment 1	26	78	22				
Experiment name							
Words from list		52	48				
Ellwanger et al. (1999)				18	100	0	March to sample task. Difficult condition only.
(Weighed) average	122	79.9	20.1	93	97.4	2.6	From Ellwanger et al. (1996), the mean of the 'experimenter name' and 'words from list' condition is included in the average.
Real life probes							
Farwell & Donchin (1991)							
Experiment 2	4	100	0	4	75	0	
Johnson & Rosenfeld (1992)	17	76	24	14	90	10	Specificity is the mean of the two irrelevant tested.
Rosenfeld et al. (1991)							
Experiment 1	13	92	8				
Experiment 2				15	87	13	Delay was 7 - 12 days.
Delay and rehearsal	8	88	12				
Delay	8	38	62				
Farwell & Smith (2001)	3	100	0	3	100	0	
(Weighed) average	53	79.2	20.8	36	87.9	9.3	
Weighed average over all studies	515	79.9	18.5	266	92.0	6.5	

## **Chapter 4: The P300 is Sensitive to Concealed Face Recognition**

Meijer, E.H., Smulders, F.T.Y., Merckelbach, H.L.G.J., & Wolf, A. G. (2007). The P300 is sensitive to concealed face recognition. *International Journal of Psychophysiology*, 66, 231-237.

### **Abstract**

In two experiments, we investigated whether a P300 based Concealed Information Test (CIT) can be used to detect concealed face recognition. The results show that detection of concealed face recognition is highly successful when stimuli depict persons who are personally highly familiar, and instructions to conceal recognition are given. When pictures depict recognized, but personally less familiar faces, and no specific instructions to conceal recognition are given, detection is unsuccessful. These findings indicate that pictures of faces can be used in a P300 based CIT, and that mere recognition is not sufficient for successful detection of concealed information.

## CHAPTER 4

### *4.1 Introduction*

The P300 component of the Event Related Potential (ERP) has been used for the detection of concealed information since almost two decades (Farwell & Donchin, 1991; Rosenfeld et al., 1988). Although the practical use of P300 based Concealed Information Tests (CITs) has been limited so far (e.g., Harrington v. Iowa, 2000; Miyake, Mizutani, & Yamahura, 1993), many researchers in the field emphasize the method's strengths (Ben Shakhar et al., 2002; Elaad, 1998; Iacono, 2007). Accordingly, in its 2003 report, the authoritative National Research Council recommended further investigation of measurement of event-related potentials as an alternative or supplement to the polygraph (National Research Council, 2003).

The rationale underlying the P300 based CIT is that rare, meaningful stimuli elicit a P300 (Donchin & Coles, 1988; Rosenfeld, 2002). Stimuli that are meaningful to the individual, like autobiographical information, have been shown to elicit a P300 waveform when presented infrequently in a series, intermixed with irrelevant stimuli (Berlad & Pratt, 1995; Gray, Ambady, Lowenthal, & Deldin, 2004). Consequently, the P300 can be used to detect simulated amnesia for autobiographical facts (Rosenfeld, Ellwanger, & Sweet, 1995). When a crime has been committed, crime-related details are thought to be meaningful to the perpetrator, but not to an innocent suspect. Therefore, if crime-relevant details are presented infrequently in a series of equally plausible, but crime-unrelated details, the crime-related details are expected to elicit a P300, but only so in guilty suspects.

The majority of studies investigating P300 based CITs used words or short phrases as stimuli. These words or phrases generally refer to either autobiographical information (e.g., Rosenfeld et al., 1995), mock crime related information (e.g., Farwell & Donchin, 1991; Rosenfeld et al., 1988), or incidentally rehearsed information (e.g., Allen et al., 1992; Ellwanger et al., 1996). In a typical P300 based CIT, these words or phrases belong to one of three types, namely targets, probes and irrelevants (e.g., Farwell & Donchin, 1991). The probe is the stimulus of interest, of which the participant denies knowledge. The target is a stimulus that has been made task relevant through instructions, and serves to ensure participant cooperation. The irrelevant stimuli are equally plausible, but unrelated to the crime. The participant is instructed to acknowledge recognition of the target by pressing one button, and press another button with any other stimulus (probe and irrelevants). For an innocent participant, probe and irrelevants are indistinguishable, and thus no difference in P300 is expected. For a



## THE P300 AND CONCEALED FACE RECOGNITION

guilty participant, on the other hand, the probe should stand out and elicit a P300.

As noted, the majority of studies investigating P300 based CITs used words or short phrases as stimuli. Photographic stimuli have seldom been used, even though especially the use of pictures of faces might be relevant. In this study, we investigated whether pictures of faces can be used in a P300 based CIT.

Many studies have investigated the cognitive processes underlying face perception and recognition. Associated ERP components include an occipito-temporal N170, thought to reflect face perception (Bentin, Allison, Puce, Perez, & McCarthy, 1996), the inferior temporal N250r found in repetition priming paradigms (Begleiter, Porjesz, & Wang, 1995; Schweinberger, Pickering, Burton, & Kaufmann, 2002), the centro-parietal N400 component that is thought to reflect facilitation in accessing postperceptual or semantic memory codes for people (Schweinberger & Burton, 2003), and a centro-parietally distributed late positive component (P300/LPC).<sup>23</sup> Especially the latter has been found to be sensitive to the recognition of familiar faces, whether famous (Henson, Goshen-Gottstein, Ganel et al., 2003), learned during a study phase (Joyce & Kutas, 2005; Paller, Bozic, Ranganath, Grabowecy, & Yamada, 1999; Paller, Gonsalves, Grabowecy, Bozic, & Yamada, 2000; Paller, Ranganath, Gonsalves et al., 2003), or the participant's own face (Ninomiya, Onitsuka, Chen, Sato, & Tashiro, 1998). Moreover, a late positive ERP component in response to familiar faces has even been observed in prosopagnosia patients, regardless of their incapability of familiar face recognition (Bobes, Lopera, Comas et al., 2004; Renault, Signoret, Debrulle, Breton, & Bolgert, 1989). However, paradigms used in the studies cited above vary greatly and in each case deviate from a typical P300 based CIT. Still, the finding that familiar faces elicit a distinct late positive ERP component supports the idea that ERPs may serve as indicators of concealed face recognition. With these considerations in mind, the aim of our first experiment was to explore whether the P300 could be used to infer concealed face recognition.

In a second experiment, we tested whether mere recognition was sufficient to detect face recognition. In a typical CIT experiment, participants are instructed to conceal recognition of the probe. But several studies have shown that a P300 can also be elicited under passive conditions, i.e., without specific task instructions (Polich, 1987, 1989; but see

---

<sup>23</sup> Some authors have suggested that the N400 and the P300/LPC are related to the same cognitive contextual integration process. See, for example, Read, Nenov, and Halgren (1994).

Bennington & Polich, 1999), indicating that, at least to some extent, the P300 indexes automatic processing (Sommer, Leuthold, & Matt, 1998). In the second experiment, we tested whether a P300 to familiar faces is also present in the absence of specific instructions to conceal recognition of the probe.

## 4.2 Experiment 1

### 4.2.1 Methods

#### *Participants*

Participants were 24 undergraduate students (4 men) at Maastricht University (average age 23.1 years; range 18 - 35). They read and signed an informed consent, and received course credits for their participation. The experiment was approved by the ethical committee of the Faculty of Psychology, Maastricht University.

#### *Stimuli*

Every participant was asked to bring two passport photos: one of a sibling and one of a good friend. The persons on the two photos had to be of the same sex. These photos were scanned and converted to grayscale. Stimulus size was 49mm by 66mm and viewing distance was 1 meter.

#### *Experimental design and procedure*

Participants were allocated to groups of three. For each member of a group, stimulus material consisted of the two pictures they brought (target and probe) plus the four pictures the other two participants brought (irrelevants). For half the participants, the picture of their sibling served as target and the picture of their friend as probe, while for the other half the picture of their friend served as target and the picture of their sibling as probe. This way, a design was created in which every stimulus served equally often as target, probe and irrelevant, thus minimizing physical stimulus confounds (Luck, 2005).

Participants were instructed to acknowledge recognition of the target by pressing one of two buttons placed under their left and right index fingers, respectively, and pressing the other button for all unfamiliar pictures. They were explicitly instructed to deny recognition of the probe, by classifying it as unfamiliar. The assignment of the target to the left and right button was balanced over participants.

Each trial started with the presentation of a picture, which was shown until the response button was pressed, with a maximum of 2500 ms. Feedback was given if no response was given after 2500 ms ('too slow!') or if the response was incorrect ('wrong!'). The intertrial interval was 2100 ms. The task consisted of 12 practice trials that served to familiarize the



## THE P300 AND CONCEALED FACE RECOGNITION

participants with the procedure, and 432 trials that were presented in three blocks of 144, with a break in between blocks that could be terminated by the participant. Thus, probe, target and each of the four irrelevants were each presented on 72 trials.

### *Data acquisition, reduction and analysis*

EEG data were recorded from 4 midline sites (Fz, Cz, Pz, Oz) and the right mastoid (M2), using Ag/AgCl electrodes, glued to the scalp with 10-20 conductive gel. All leads were online referenced to the left mastoid (M1). Horizontal and vertical electrooculograms (EOGs) were recorded using electrodes placed laterally to both eyes as well as below and above the left eye. EEG and EOG electrode impedances were below 5 k $\Omega$  and 10 k $\Omega$ , respectively. All signals were amplified using Contact Precision Instruments amplifiers. EEG was amplified 20,000 times, EOG 4,000 times. The signal was filtered online (0.1 to 30 Hz bandpass), and digitized at 200 Hz. All leads were off line re-referenced to an average of M1 and M2. Eye blink artifacts were reduced using a regression based method (Semlitsch, Anderer, Schuster, & Presslich, 1986) performed on the continuous data. After this, epochs were extracted from the continuous data, lasting from 100 ms before until 1200 ms after stimulus onset. To ensure a reliable artifact rejection, these epochs were baseline corrected, after which all trials containing amplitudes exceeding  $\pm 75 \mu\text{V}$  and all trials with an incorrect behavioral response were removed. Remaining trials were then baseline corrected on the pre-stimulus interval, and averaged per stimulus type. All 'irrelevant' trials were pooled into one average.

P300 was measured using a typical base-peak method, as well as the peak-peak method described by Rosenfeld (e.g., Rosenfeld et al., 2006). Firstly, the maximal positive 100 ms segment average was determined in the 300 to 800 ms window. This was defined as the base-peak P300 amplitude. Next, the maximal negative 100 ms segment average following this positive segment was determined. Peak-peak P300 amplitude was defined as the difference between these two segments. It has repeatedly been shown that this peak-peak method results in better detection of deception than a typical base-peak measure (e.g., Soskins, Rosenfeld, & Niendam, 2001). Therefore, this peak-peak P300 measure was used as the dependent variable in an analysis of variance. As P300 is generally largest at Pz, we limited our analysis to this site.

Apart from the analysis at group level, it is also of interest to look at the success of detecting concealed face recognition within each individual. To compute hit rates, we used a bootstrapping approach (see e.g., Wasserman & Bockenholt, 1989). This approach generates a distribution for



## CHAPTER 4

each stimulus type within each individual, allowing for statistical testing. Basically, the bootstrapping approach entails generating multiple different averages from the same set of stimuli. Suppose that, after discarding artifacts and incorrect behavioral responses, for a given participant,  $n$  probe trials and  $i$  irrelevant trials remain. Instead of averaging across all  $n$  probe trials, step 1 of the bootstrapping procedure entails drawing  $n$  trials randomly with replacement from the  $n$  probe trials, and averaging these. Step two entails scoring the P300 on this average using the P300 scoring method described above, yielding a single probe P300 score. Step one and two are then repeated for the irrelevant stimulus type, drawing  $i$  trials, yielding a single irrelevant P300 score. The third step consists of subtracting this irrelevant P300 score from the probe P300, resulting in a single difference score. The three steps are repeated 100 times, and due to the random drawing with replacement, each average consists of a different set of trials, yielding 100 separate difference scores.

To state with 95% confidence that probe P300 is larger than irrelevant P300, and thus that the detection of concealed information was successful, we checked whether the mean difference score minus  $1.65 \times$  the standard deviation was greater than zero. If so, the probe was classified as recognized. In order to compute the hit rate based on RTs, an identical procedure was applied to the behavioral data.

### 4.2.2 Results and discussion

#### *Behavioral data*

Participants made, on average, 8.4 % errors (incorrect or no button press) on target trials, 1.2 % errors on probe trials, and 0.12 % errors on irrelevant trials. A  $2 \times 2$  ANOVA on these error rates with stimulus type (probe, irrelevant) as repeated measure and probe type (sibling, friend) as between-subjects factor revealed no significant interaction [ $F(1,22) = 1.76, p = .20$ ], indicating that the probe-irrelevant difference did not depend on probe type. The factor probe type was therefore dropped from the analysis. The subsequent ANOVA showed that participants indeed made more errors on probe than on irrelevant trials [ $F(1,23) = 9.8, p < .01$ ]. Table 4.1 shows the RTs for the different stimulus types. A  $2 \times 2$  ANOVA on these RTs with stimulus type and probe type as factors again revealed no significant interaction [ $F(1,22) = 2.27, p = .15$ ], and the factor probe type was therefore dropped from the analysis. The subsequent ANOVA yielded a significant effect [ $F(1,23) = 45.2, p < .001$ ], showing that reactions to probes were slower than those to irrelevants. Bootstrapping of the reaction times resulted in a hit rate of 79%, i.e., for 19 out of 24 participants

## THE P300 AND CONCEALED FACE RECOGNITION

**Table 4.1.** Mean peak to peak P300 amplitude in uV (P300) and mean reaction times in ms (RT) to the different stimulus types (target, probe, and irrelevant) and number of hits in Experiment 1 (personally highly familiar probe with instructions to conceal recognition: left panel) and Experiment 2 (personally less familiar probe without instructions to conceal recognition: right panel). Standard deviations are given between parentheses.

	Experiment 1		Experiment 2	
	P300	RT	P300	RT
Target	23.1 (6.6)	586.9 (54.3)	25.5 (7.3)	626.9 (61.5)
Probe	16.7 (6.5)	555.4 (57.8)	11.4 (6.1)	556.5 (60.2)
Irrelevant	9.6 (5.8)	502.3 (46.0)	11.2 (5.6)	563.0 (49.6)
Hits	22/24 (92%)	19/24 (79%)	4/24 (17%)	1/24 (4%)

concealed face recognition was detectable through their slower responses on probe stimuli.

### *P300 data*

Figure 4.1 (left panel) shows the grand average ERP waveforms at all electrode sites. A 2 X 2 ANOVA on peak-peak P300 amplitude at Pz with stimulus type and probe type as factors revealed no significant interaction [ $F(1,22) < 1$ ], and the factor probe type was therefore dropped from the analysis. The subsequent ANOVA revealed a main effect for stimulus type [ $F(1,23) = 54.4$ ,  $p < .001$ ], indicating that the P300 elicited by probes was larger than the P300 elicited by irrelevants.<sup>24</sup> 4.1 presents mean P300 amplitude and individual classification. The hit rate for the peak-peak P300 measurement was 92%, i.e., 22 out of 24 participants displayed a P300 for probes that sufficiently deviated from that for irrelevants to be of diagnostic value. The peak-peak measure outperformed a base-peak measure, which achieved a hit rate of 71% (17 out of 24 participants). These results show that, whether based on RTs or P300, our face recognition based CIT was highly successful.

To test whether this P300 associated with concealed face recognition can be explained by mere recognition processes, we conducted a second experiment. To isolate recognition processes, no instructions to conceal recognition of the probe were given. Furthermore, we addressed a point of concern, namely that the stimuli were photographs acquired through the participants themselves. As a consequence, targets and probes may have been recognized as familiar faces, but also as their self-brought pictures. If so, the P300 found might represent photo recognition instead of

<sup>24</sup> Analysis of the base-peak P300 measure yielded similar results (main effect for stimulus type [ $F(1,23) = 45.0$ ,  $p < .001$ ]).



## CHAPTER 4

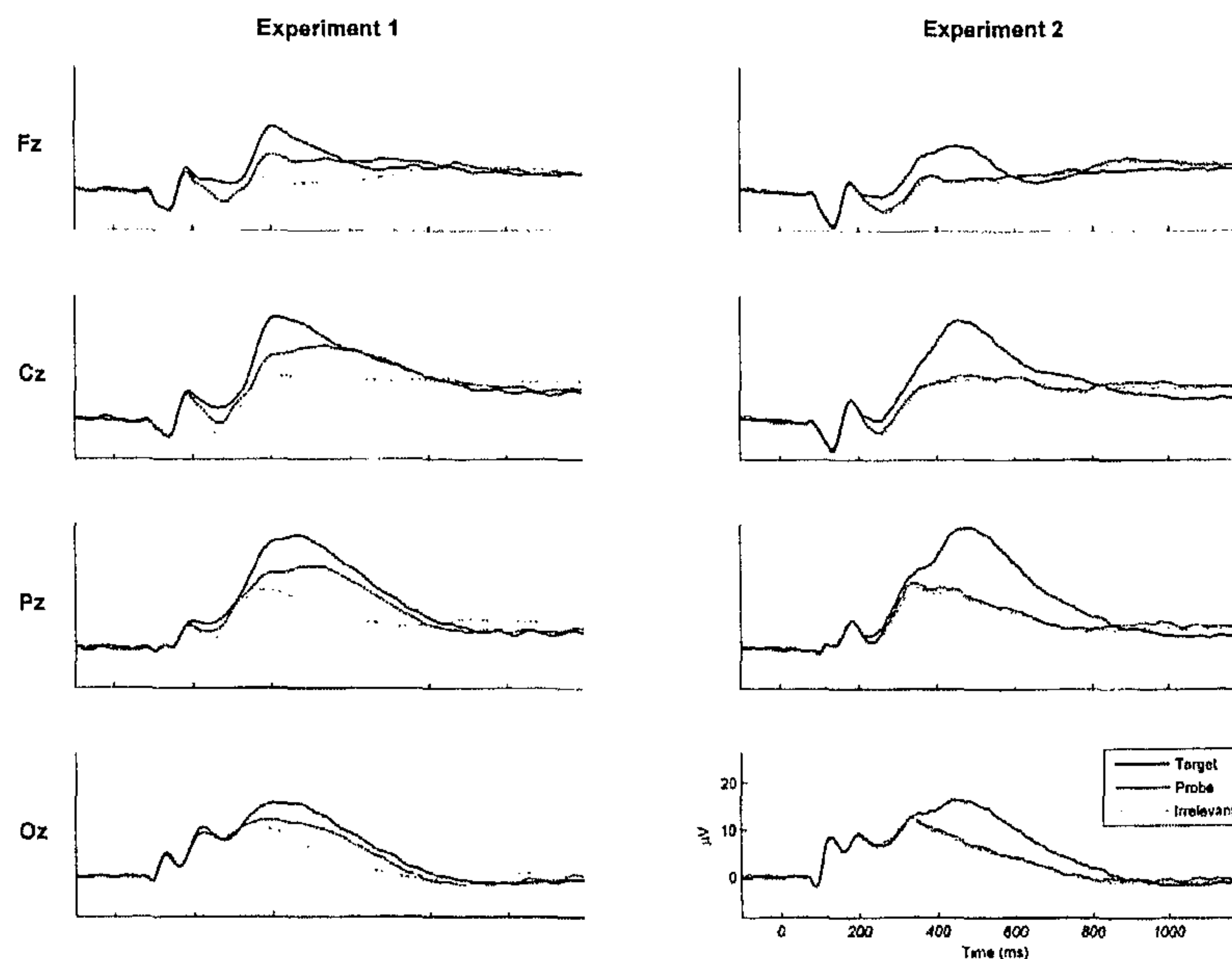


Figure 1: Grand average ERPs at Fz, Cz, Pz, and Oz as a function of stimulus type (target, probe, and irrelevant) in Experiment 1 (personally highly familiar probe with instructions to conceal recognition: left panel) and Experiment 2 (personally less familiar probe without instructions to conceal recognition: right panel).

face recognition per se. In the second experiment, we used as stimuli pictures of university professors who had lectured in courses that were attended or unattended (in two groups) by the participants.

### 4.3 Experiment 2

#### 4.3.1 Methods

##### *Participants*

Participants were 29 first-year students at Maastricht University. Five of them were dropped from the analysis because they could not correctly identify the critical photos after the experiment (see experimental design and procedure section). Thus, the remaining sample consisted of 24 participants (average age 19.7 years; range 18 – 24, 2 men). Twelve of them were first-year psychology students, whereas the other 12 were first-year health sciences students. They read and signed an informed consent, and received course credits for their participation. The experiment was



## THE P300 AND CONCEALED FACE RECOGNITION

approved by the ethical committee of the Faculty of Psychology, Maastricht University.

### *Stimuli*

Photos of the faces of six university professors served as stimuli. Two of these professors taught at the Faculty of Psychology, two at the Faculty of Health Sciences, and two at the Faculty of Law. All six professors had taught a first-year course that year which had already been completed. As a consequence, two faces were familiar to each participant, while four were not. Twelve pictures were taken of each professor, four from a slight angle from the left, four from *en face*, and four from a slight right angle. This was done to reduce the chance that any physical property of the picture would lead to an easy classification. Size and resolution of the pictures were identical to those in Experiment 1.

### *Experimental design and procedure*

For each participant, one of the two familiar faces was denoted as the target. Participants were instructed to rapidly press one button upon presentation of the face of the professor teaching course X, and the other button upon presentation of any other face. In contrast to Experiment 1, no specific reference was made to the fact that, besides the target, another familiar face would be presented, nor to concealing recognition of this face. Participants were then shown a picture of the professor teaching course X. Which of the two familiar faces served as the target was balanced over participants. The other familiar face served as probe. The remainder of the procedure was identical to Experiment 1, with one exception: stimuli were presented in six blocks containing 72 stimuli each (i.e., all 12 pictures of each of the 6 professors presented once), preceded by a practice block that also contained 72 stimuli. After the ERP recording, all participants were given a pencil-and-paper forced-choice recognition test. This test consisted of a form on which participants had to link the pictures to a specific course, or check the 'unknown' option if they had not seen the person prior to the experiment. This test was to make sure that both the target and the probe, but not the irrelevants, had indeed been recognized.

### *Data acquisition, reduction and analysis*

Data acquisition and analysis were identical to Experiment 1, with the exception that, due to technical problems, electrodes were rereferenced to M2. Only the trials with pictures of professors who taught at the psychology or health sciences faculty were analyzed. This was done because only psychology and health sciences students participated, and in this way all pictures that were included in the averages served as targets, probes, and irrelevants.

## CHAPTER 4

### 4.3.2 Results

#### *Behavioral data*

Participants made on average 5.8 % errors to targets, 0.12 % errors to probes, and 0.20 % errors to the irrelevant stimuli. A 2 X 2 ANOVA on these error rates with stimulus type (probe, irrelevant) and study (psychology, health sciences) as factors revealed no interaction effect [ $F(1,22) < 1$ ]. The factor study was therefore dropped from the analysis. The subsequent ANOVA showed that the difference between the number of errors to probes and irrelevants was not significant [ $F(1,23) = 1.56$ ,  $p = .22$ ]. Table 4.1 shows the RTs to the different stimulus types. A 2 X 2 ANOVA on these RTs with stimulus type and study as factors revealed no interaction [ $F(1,22) < 1$ ], and the factor study was therefore dropped from the analysis. The subsequent ANOVA did not reveal an effect for stimulus type [ $F(1,23) = 2.23$ ,  $p = .15$ ]. Bootstrapping of the reaction times resulted in successful detection of recognition in only one participant (hit rate 4%).

#### *P300 data*

Figure 4.1 (right panel) presents grand average waveforms of Experiment 2. A 2 X 2 ANOVA on peak-peak P300 amplitude at Pz with stimulus type and study as factors revealed no interaction [ $F(1,22) = 1.47$ ,  $p = .24$ ], and the factor study was therefore dropped from the analysis. The subsequent ANOVA did not reveal an effect of stimulus type [ $F(1,23) < 1.0$ ].<sup>25</sup> Table 4.1 gives the mean P300 amplitude and hit rate. Both the peak-peak and the base-peak method correctly classified 17% of all participants, which means that in 4 participants the P300 amplitude for probes deviated sufficiently from that for irrelevants to be of diagnostic value.

#### *Comparison of Experiments*

Including the data from both experiments in one 2 X 2 ANOVA with Experiment (Experiment 1, Experiment 2) and stimulus type (probe, irrelevant) as factors indeed showed a strong interaction [ $F(1,46) = 42.1$ ,  $p < .001$ ], confirming the difference in the P300's sensitivity to probes between the two experiments. Similarly, the RT data showed a strong interaction between experiment and stimulus type [ $F(1,46) = 43.7$ ,  $p < .001$ ].

## 4.4 General discussion

The aim of this study was two-fold. Firstly, we investigated whether a typical P300 based CIT can be used to detect concealed face recognition. Secondly, we investigated whether this effect could be explained by mere

---

<sup>25</sup> Analysis of the base-peak P300 measure yielded a borderline significant main effect for stimulus type ( $[F(1,23) = 4.0$ ,  $p = .06]$ ).



## THE P300 AND CONCEALED FACE RECOGNITION

recognition processes. The CIT in Experiment 1 was successful, both in terms of statistical significance at the group level as well as in terms of individual classification. Experiment 2 showed that this effect could not be explained by mere recognition.

The results from Experiment 1 support the use of pictures of faces in a P300 based CIT. This may help to identify someone as a member of a terrorist network or a criminal organization, or may help to identify a rapist by presenting the face of the victim. The latter may be especially useful when the rapist does not know the name of the victim. If this pattern generalizes to other pictures, the range of items available for forensic CITs may be increased, thereby helping to overcome a practical problem often raised for the CIT, namely that in real crimes insufficient numbers of items are available (Podlesney, 1993, 2003).

At the same time, the results from Experiment 2 show that mere recognition is insufficient for a successful CIT. Most likely, because of the instructions to deny recognition of the probe in Experiment 1, probe and irrelevants were processed differently. In Experiment 2, the absence of specific instructions may have resulted in similar processing of probe and irrelevants, and thus similar P300s. This is consistent with a view that P300 in response to the probe in Experiment 1 was due to intentional rather than to automatic processing, and concurs with previous studies showing an increased P300 with specific task instructions (Castro & Diaz, 2001; Rosenfeld, Soskins, Bosh, & Ryan, 2004). Furthermore, the large P300 to the target stimulus in Experiment 2 corroborates the P300 eliciting effect of task instructions, as this stimulus was equally familiar as the probe, but was made task relevant through instructions.

Another difference between the two experiments was that the pictures of professors used in Experiment 2 were personally less familiar than the pictures of siblings and good friends used in Experiment 1. Emotionally significant stimuli have been shown to elicit a P300 (Ito & Cacioppo, 2000; Schupp, Cuthbert, Bradley et al., 2000) and this may be (partly) responsible for the P300 in Experiment 1. This explanation would be in line with recent studies by Rosenfeld and colleagues (2006, 2007), who found that detection of well-rehearsed, but personally non-significant word stimuli (the name of the experimenter) was inferior to that of personally significant stimuli (participants' own name), and would also be consistent with recent models of face recognition (Breen, Caine, & Coltheart, 2000; Ellis & Lewis, 2001). According to these models, brain systems responsible for the processing of faces include a cognitive route providing access to semantic and biographical information about the faces. Additionally, a



## CHAPTER 4

second, separate route is responsible for the generation of an affective response to familiar faces. Face recognition units are activated when a familiar face is recognized, and subsequently, one or both of these routes can be triggered. Given the relationship between the participants and the persons on the pictures, the faces in Experiment 1 may have triggered the affective route, whereas the faces in Experiment 2 have not. A larger P300 to faces that trigger the affective route would, in turn, be in line with models of P300 that emphasize the importance of stimulus meaning (Johnson, 1993).

Two limitations of the current experiments deserve some comments. Firstly, as mentioned, the stimuli used in Experiment 1 were photographs acquired through the participants themselves. As a consequence, targets and probes may have been recognized as familiar faces, but also as their self-brought pictures. If so, the P300 found in Experiment 1 might represent photo recognition instead of face recognition per se. Secondly, two factors were modified between Experiment 1 and Experiment 2. In Experiment 2, task instructions did not make any specific reference to presentation of the probe, nor to concealing its recognition, and stimuli depicted someone personally less familiar to the participants. Either of these factors may have affected P300 amplitude. Whether one of these factors alone is sufficient for a successful P300 based CIT should be clarified by future studies.

In sum, the present results demonstrate that pictures of faces can be used effectively in a P300 based CIT, and that mere recognition is not sufficient for successful detection of concealed information.

## **Chapter 5: Psychopathic traits and the detection of concealed information: Event related potential and skin conductance combined**

Meijer, E.H., Smulders, F.T.Y., Merckelbach, H.L.G.J., & Elbers, N. (in prep.) Psychopathic traits and the detection of concealed information: Event related potential and skin conductance combined.

### **Abstract**

The present paper examines the association between psychopathic traits, skin conductance response (SCR), and the P300 component of the EEG in a concealed information test (CIT). Skin conductance and EEG were measured simultaneously, while stimuli were presented with a short inter stimulus interval (ISI). To overcome the problem of overlapping SCRs with these ISIs, stimuli were presented in a specially balanced order. Results showed that detection of concealed information with P300 recovery, but not with P300 amplitude nor with SCR correlated with psychopathic traits. Furthermore, SCR did not correlate with P300. To the extent that psychopathy moderates autonomic responding in a CIT, thereby reducing its diagnostic accuracy, measurement of the P300 may be a solution

## CHAPTER 5

### *5.1 Introduction*

The polygraph has been around in criminal investigations since the 1930s (Larson, 1932; Marston, 1938). Its use has, however, been subjected to a heated debate with the Control Question Technique (CQT) capturing most of the criticisms (see e.g., Ben Shakhar, 2002; Fiedler et al., 2002; Lykken, 1998; National Research Council, 2003). During a CQT, physiological responses to relevant questions like “On December 12<sup>th</sup>, did you shoot John Doe?” are compared with responses to control questions, e.g., “During the last 20 years, did you ever hurt someone?” Stronger responding to relevant questions is thought to be indicative of guilt, whereas stronger responding to control questions is said to indicate innocence. The CQT has been criticized for a number of reasons, including lack of standardization, lack of scientific underpinning, and high risk of false positive outcomes (Ben Shakhar, 2002; Lykken, 1998; National Research Council, 2003). A different polygraph technique is the Guilty Knowledge Test (GKT; Lykken, 1959, 1960), also referred to as the Concealed Information Test (CIT). Typically, CIT questions deal with crime related information presumably only known to the police and the perpetrator. Multiple answer alternatives are presented serially (e.g., the victim was stabbed with (a) an ice pick (b) a letter opener (c) a pair of scissors (d) a tent peg (e) a dagger), while a range of psychophysiological parameters are recorded. Consistent stronger physiological responding to the correct alternative across multiple questions is thought to be indicative of crime related knowledge, and from this involvement in the crime under investigation can be inferred.

A number of physiological measures have been shown to be successful in the detection of concealed information. Among the physiological measures that have received the most attention are skin conductance response (SCR) and the P300 component of the event related potential (ERP). The robustness of SCR in discriminating guilty from innocent participants in CIT paradigms has repeatedly been shown. In a quantitative meta-analysis of 22 studies, MacLaren (2001) found that the proportion of correctly classified guilty participants (i.e., sensitivity) was 76% and the proportion of correctly classified innocent participants (i.e., specificity) was 83%. A more recent meta-analysis (Ben Shakhar & Elaad, 2003) found an average effect size ( $d$ ) of 1.55 in 80 studies. This effect size was even larger (3.12) for a subset of studies that used optimal conditions like motivational instructions, a deceptive verbal response made by the participants, and a CIT consisting of at least 5 questions.



One potential problem with the use of skin conductance in the detection of concealed information, however, is that its detection accuracy might depend on individual differences. In this regard, the psychopathy dimension is particularly relevant. Core features of this condition include impulsive and antisocial behavior, deceit and superficial charm, and a lack of empathy and remorse. The incidence of psychopathy is estimated at 15%-25% in US inmates, and a little under 1% in the community (Blair, Mitchel, & Blair, 2005; Hare, 1999). Meanwhile, autonomic underarousal and hyporesponsiveness are prominent features of psychopathy (Lorber, 2004). This hyporesponsivity may in turn undermine CIT accuracy. If so, ironically, individuals prone to criminal behavior have a relatively low probability of being detected. Indeed, Verschuere, Crombez, Declercq and Koster (2005) found that prisoners who scored high on certain psychopathy subscales exhibit both a reduced overall SCR as well as reduced differential responding to the correct and incorrect alternatives in a CIT. This effect was, however, not replicated in a more recent study (Verschuere, Crombez, Koster, & De Clercq, 2007), meaning that the evidence for the undermining effects of psychopathy on the CIT remains inconclusive.

To the extent that psychopathic traits indeed moderate autonomic responding in a CIT, P300 measurement may offer a solution. This measure was first used in CITs during the late 80's (Farwell & Donchin, 1991; Rosenfeld et al., 1988), and received considerable attention since. The P300 wave is sensitive to a wide variety of cognitive functions, including working memory and context updating (Donchin & Coles, 1988), and recognition memory (Friedman & Johnson, 2000; Woodruff, Hayama, & Rugg, 2006; Yonelinas, 2002). It is highly sensitive to rarity and (task) relevance (Donchin & Coles, 1988), and can thus be used in a CIT. The rationale behind this is that when crime-relevant details are presented in a series intermixed with crime-irrelevant, but equally plausible details, the crime-relevant details elicit a P300 only in guilty participants. Research has consistently shown that the P300 performs well in the detection of concealed information. The majority of studies, using either the mock crime paradigm, autobiographical stimuli or learned material as stimuli, yielded sensitivities typically exceeding 80 or even 90% (Allen & Iacono, 1997; Allen et al., 1992; Farwell & Donchin, 1991; Miller & Rosenfeld, 2004; Rosenfeld et al., 1995; Rosenfeld et al., 1999; Rosenfeld et al., 2004).

As noted, the P300 is sensitive to a wide variety of cognitive functions. The extent to which psychopathy moderates these functions is largely unknown. Findings from studies looking at P300 in CIT-related tasks are mixed. Jutai, Hare, and Connolly (1987) found no difference between

## CHAPTER 5

psychopaths and controls in an auditory oddball task using phonemes as stimuli. Raine and Venables (1988), on the other hand, found an increased P300 amplitude and P300 recovery time to target stimuli in psychopaths using a visual oddball task with digits as stimuli. This led the authors to conclude that psychopaths might have an enhanced ability to attend to task relevant events (Raine, 1989; Raine & Venables, 1988). More recent studies, however, found reduced P300 to target stimuli in psychopaths in both visual (Kiehl, Hare, Liddle, & McDonald, 1999) and auditory (Kiehl, Bates, Laurens, Hare, & Liddle, 2006) oddball tasks. In all, the evidence regarding the effect of psychopathy on P300 remains inconclusive, and suggests that psychopathic individuals' P300 may be increased in certain conditions, but decreased in others. All studies cited above relied on diagnosed psychopaths derived from populations of inmates. Psychopathic traits, however, are not limited to incarcerated offenders (Babiak & Hare, 2006), and can be assessed in noncriminal populations using the Psychopathic Personality Inventory (PPI; Lilienfeld & Andrews, 1996). In a first attempt, Miller and Rosenfeld (2004) used only a limited number of items from the PPI, and found no difference in P300 based detection rates between high and low scorers.

In all, the effects of psychopathy on P300 and SCR during a CIT procedure remain largely unclear. Relevant research, however, is complicated by the fact that P300 and SCR are typically recorded on widely different time scales. SCR has an onset latency and rise time of 1 to 3 s, and a half recovery time of up to 10 s (Dawson, Schell, & Filion, 2000). Because of this recovery time, it is usually recorded with an inter stimulus interval (ISI) of 20 s or more. Latency of the P300 on the other hand, is less than 1 s, and in typical P300 oddball tasks, stimuli are presented with an ISI between 1 and 3 s. Longer ISIs result in a decreased P300 (Polich, 1990). Measuring skin conductance data at such short ISIs results in overlapping responses, which is why concurrent measurement of SCR and ERP has received little attention (for exceptions see Bahramali, Gordon, Lim et al., 1997; Barry, Feldmann, Gordon, Cocker, & Rennie, 1993; Lim, Gordon, Rennie et al., 1999). We report results of an experiment in which ERPs and SCRs were measured simultaneously during a CIT, using a short ISI. To overcome the problem of strongly overlapping SCRs, stimuli were presented in a specially balanced order.

### **5.2 Method**

#### *Participants*



## COMBINING SCR AND SVT IN A CIT

Participants were 30 undergraduate students (11 men; mean age 21.3; range 19-28) from Maastricht University. All participants read and signed a consent form and received course credits for their participation. The experiment was approved by the standing ethical committee of the Faculty of Psychology.

### *Materials*

*PPI.* To assess psychopathic traits, the Dutch translation of the Psychopathic Personality Inventory (PPI; Jelicic, Merckelbach, Timmermans, & Candel, 2004; Lilienfeld & Andrews, 1996) was used. Following Benning, Patrick, Hicks, Blonigen, and Krueger (2003) and Verschuere et al. (2005), we calculated the Fearless Dominance factor and the Impulsive Antisocial factor by summing scores across the appropriate subscales while compensating for the fact that these subscales consist of different numbers of items.

*SCR.* Skin conductance was measured using a 24 bit DC 0.5 Volt system. Two Beckmann silver/silver chloride (Ag-AgCl) electrodes (5 mm in diameter) were placed on the medial phalanges of the first and second fingers of participants' non-dominant hand. Electrodes were filled with isotonic electrode paste (0.9% NaCl). Respiration was measured using a strain gauge attached around the thorax. All data were acquired using Contact Precision Instruments with a sample rate of 200 Hz.

*EEG.* EEG data were recorded from 4 midline sites (Fz, Cz, Pz, Oz) and the right mastoid (M2), using Ag/AgCl electrodes, mounted to the participants head using 10-20 conductive gel. All leads were online referenced to the left mastoid (M1). To remove eye movement artifacts from the EEG, horizontal and vertical eye movement (EOG) were recorded using electrodes placed at the outer canthi of both eyes as well as below and above the left eye. All EEG electrode impedances were below 5 k $\Omega$ , all EOG electrode impedances below 10 k $\Omega$ . All channels were amplified using Contact Precision Instruments amplifiers. EEG was amplified 20k times, EOG 4k times. The signal was filtered online with a bandpass from 0.1 to 30 Hz. and digitized at 200 Hz.

### *Stimuli and task*

In a typical P300 based CIT, three types of stimuli are presented: targets, probes, and irrelevant (e.g., Farwell & Donchin, 1991). The probe is the stimulus of interest, of which the participant denies recognition (e.g., the true murder weapon). The target is a stimulus that has been made task relevant through instructions, and serves to ensure stimulus processing by the participant. The irrelevant stimuli are plausible, but incorrect alternatives.

## CHAPTER 5

To overcome the problem of overlapping SCRs associated with an ISI shorter than the signal's recovery time, stimuli were presented in a balanced order using M-sequences (Buracas & Boynton, 2002). Characteristic of an M-sequence is that each stimulus type is preceded by every other stimulus type equally often, up to the  $n^{\text{th}}$  order. Consequently, averaging per stimulus category results in averages that all have an identical 'history' (see Appendix A for an example of a 3<sup>rd</sup> order balanced sequence with three stimulus types). This means that, although the average response for each type is biased by responses to previous stimuli, this bias is identical for every stimulus type. Therefore, the differences between the averaged responses to each stimulus type are unbiased.

In addition to the probe, target, and irrelevant stimulus type, a null event was also presented. Here, no stimulus was presented, but the time point was marked. By averaging time locked to this time point, we obtained an estimate of the bias produced by the responses to previous stimuli. Subtracting this estimate from the average response to a stimulus type resulted in an unbiased average response to that stimulus type.

The task consisted of 6 blocks, with each block representing a single question referring to a detail of the mock crime (see procedure). The order of these blocks was determined by means of a Williams latin square (Williams, 1949). All answer alternatives presented within one block belonged to one of 8 types: the target, probe, 5 irrelevants and the null event. Within a block, each alternative was presented 8 times. Thus, stimuli were presented in a series of 63 ( $8 * 8$ ), with the one missing trial being a consequence of the mathematical properties of the M-sequence. Twelve leading and 12 trailing trials were added to the sequence of 63, creating a 'history' and 'future' for the initial and last stimuli. Stimuli were balanced up to the 8<sup>th</sup> order, and presented with an ISI that varied from 2250 to 2750 ms, creating a bias free interval of approximately 20 s. Participants were instructed to acknowledge recognition of the target by pressing a button labeled 'yes' while pressing the button labeled 'no' for all non-target stimuli.<sup>26</sup> The assignment of 'yes' and 'no' to the left and right button was balanced over participants.

### *Procedure*

After reading and signing the informed consent, participants filled out the PPI. Participants were then asked to pick one of two envelopes which both

---

<sup>26</sup> For example; the perpetrator left behind a coffee mug at the crime scene. For the block pertaining to the object that was left behind by the perpetrator, the answer 'umbrella' was denoted as target, and instructions were: "please press the yes button if the answer is 'umbrella', and no for all other answers."



## COMBINING SCR AND SVT IN A CIT

contained the same instructions to commit a mock crime. The mock crime consisted of stealing €10 from a red wallet in a room within the university building. Besides the amount of money stolen, the color of the wallet and the room number, CIT questions addressed the image on a poster hanging inside the room, the object found at the crime scene and the object left by the perpetrator. The cover-up story of the two envelopes aimed at producing the belief that the experimenter was unaware of the task completed by the participant. After committing the mock crime, the participants returned to the laboratory. They were then informed about the CIT, and were promised a €2.50 reward for an innocent outcome. Next, EEG and SCR electrodes were attached, and testing began. Testing took place in a dimly lit, soundproof chamber, and participants were monitored from a control room by means of a video camera. After the test, participants filled out a forced choice test to make sure they indeed correctly recalled the crime related details.

### *Data reduction and analysis*

*SCR.* Within each block, the 12 leading and 12 trailing questions were excluded from the analysis. The remaining 63 trials were epoched from -1 to 20 s relative to stimulus onset. These epochs were baseline corrected at the sample preceding stimulus onset and then averaged per stimulus type, resulting in 8 averages. The average of the null event was subtracted from the average of each stimulus type. Subsequently, the maximum positive deflection in the 1 to 5 s post stimulus window was defined as the SCR. Finally, within-block standardized scores were computed by subtracting the mean of 6 responses (5 irrelevants and 1 probe) from the response to the probe item and dividing that by the standard deviation of all six responses. This procedure serves two purposes. First, it eliminates individual differences in responsivity (Ben-Shakhar, 1985). Second, the resulting parameter is a measure for the detection of concealed information. If the response to the probe is similar to that to the irrelevant stimuli, the standardized score will approach 0. Stronger responding to the probe than to the irrelevant stimuli will result in a positive z-score, and a significant positive deviation from 0 is indicative of successful detection of concealed information. The steps described above were repeated for each block, yielding 6 z-scores.

*EEG.* All leads were off line re-referenced to an average of M1 and M2. Eye movement artifacts were reduced using a regression based method described by Semlitsch, Anderer, Schuster, and Presslich (1986). Next, within each block, the 63 epochs lasting from 100 ms before until 1200 ms after stimulus onset were extracted from the continuous data. The average

## CHAPTER 5

amplitude during the pre-stimulus interval was subtracted, and epochs were averaged per stimulus type. P300 amplitude was measured using the peak-peak method described by Rosenfeld (e.g., Rosenfeld et al., 2006). First, the maximal positive 100 ms segment average was determined in the 300 to 800 ms window. Next, the maximal negative 100 ms segment average following this positive segment was determined. P300 amplitude was defined as the difference between these two segments. It has repeatedly been shown that this peak-peak method results in better deception detection than a typical base-peak measure (e.g., Soskins et al., 2001; see also chapter 4). Based on visual inspection of the data, a second measure of P300, namely P300 recovery was defined as the area under the curve in the 500 to 800 ms window. As P300 is generally largest at Pz, we limited our analysis to this site. Similar to the SCR data, within block standardized scores representing the probe-irrelevant difference were calculated for P300 amplitude and P300 recovery. These standardized scores are referred to as z-SCR, z-P300 amplitude and z-P300 recovery.

### *Statistical analysis*

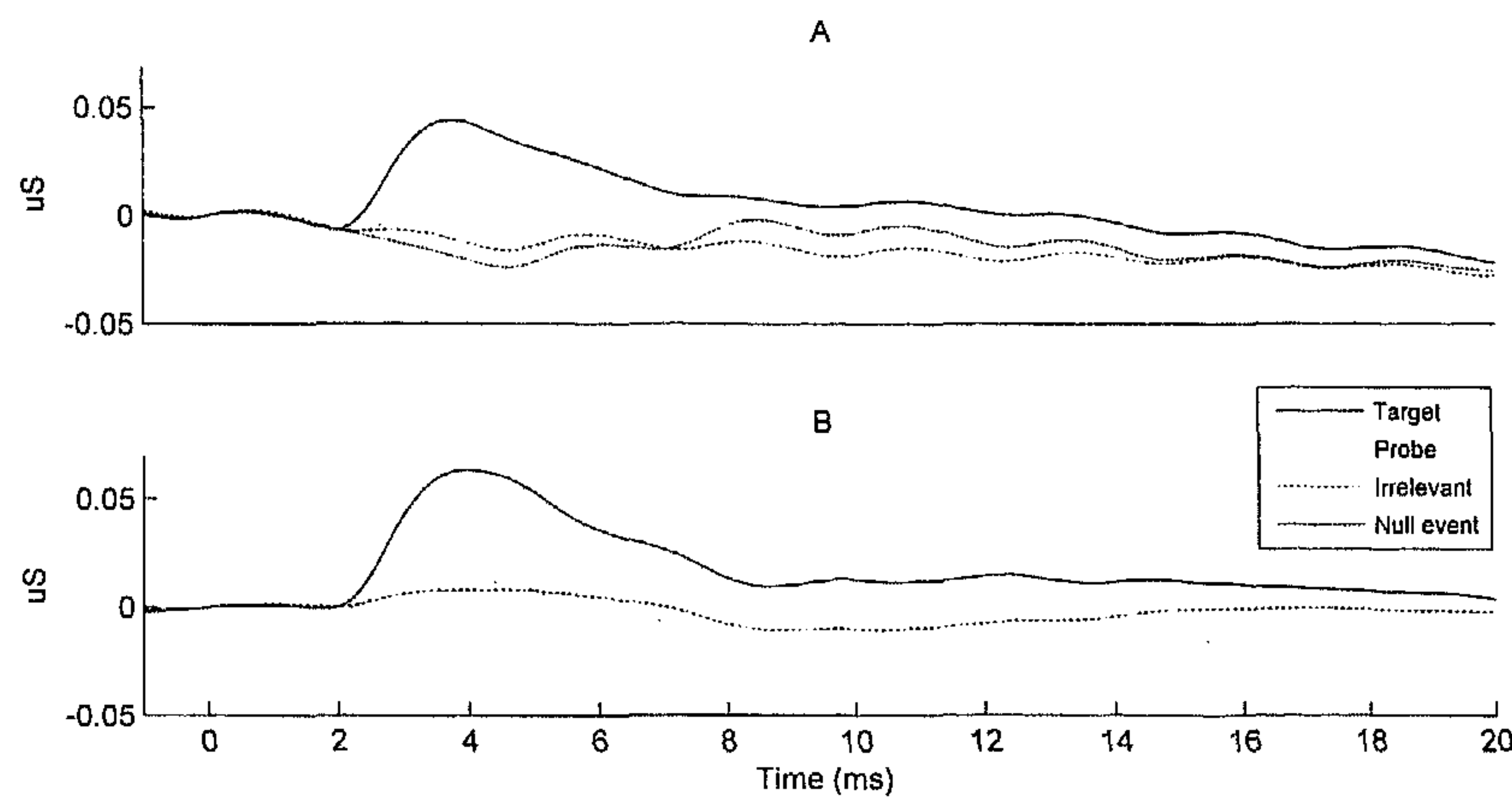
Detection efficiency is expressed using signal detection theory (SDT; National Research Council, 2003; Wickens, 2002). This method defines detection efficiency in terms of the degree of separation between the distributions of the detection measures (i.e., z-SCR, z-P300 amplitude and z-P300 recovery) for the innocent and the guilty condition. As no innocent participants were tested in the current experiment, we simulated a group of innocent participants by randomly drawing 6 values from the standard normal distribution (representing 6 responses, one to the probe and five to the irrelevant stimuli), standardizing one of these values relative to the mean and standard deviation of all 6 values, and repeating this procedure 6 times to represent 6 questions (see chapter 6). Subsequently, the distance between the centers of the simulated distribution of the innocent and the distribution of the guilty participants was computed in terms of standard deviation ( $d$ ), and the area under the Receiver Operating Characteristic (ROC) curve ( $a$ ) was computed.

### **5.3 Results**

Figure 5.1A shows the time locked SCR, averaged over all blocks, uncorrected for overlap. This figure shows in the typical SCR latency window of 1 to 5 s, a large response to the target, a smaller response to the probe, a still smaller response to the irrelevant, and no response to the null-event. The overlapping responses of other stimuli are observable as fluctuations with a period of 2.5 s, exactly the ISI. Overlap corrected



## COMBINING SCR AND SVT IN A CIT



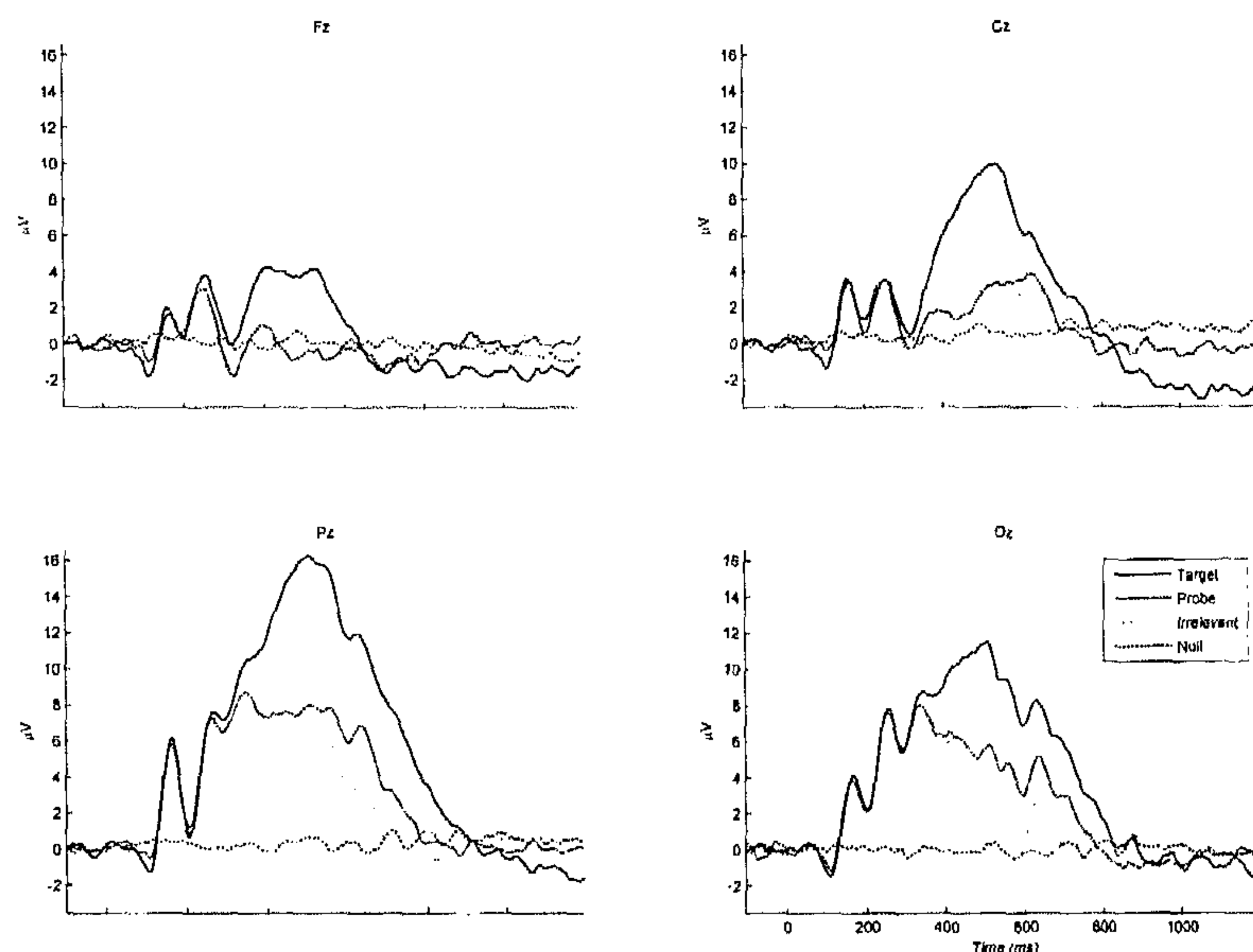
**Figure 5.1.** (A) Average skin conductance response as a function of stimulus type (target: recognized, admitted; probe: recognized, concealed; irrelevant: unrecognized; and null event: no stimulus presented) and (B) of the target, probe and irrelevant stimulus type with the null event subtracted.

waveforms were created by subtracting the response to the null-event from the other stimuli, point by point. The resulting waveforms are presented in Figure 5.1B. It is also noteworthy that although the morphology of the waveform corresponds to a typical SCR waveform, the amplitude is much smaller than the  $0.2 - 1.0 \mu\text{S}$  typically observed with long ISIs (Stern, Ray, & Quigley, 2001).

The average ERP waveforms are presented in Figure 5.2. Repeated measure MANOVAs on the z-scores (representing the probe-irrelevant difference) with block (6 levels) as a repeated measure revealed no significant linear effects for z-SCR, z-P300 amplitude or z-P300 recovery (all  $F_s < 2$ ),<sup>27</sup> meaning that differential responding did not change over time. A significant intercept was found for all three measures (z-SCR:  $F(1,29) = 6.8$ ,  $p = .01$ ; z-P300 amplitude:  $F(1,29) = 5.1$ ,  $p = .03$ ; z-P300 recovery  $F(1,29) = 28.4$ ,  $p < .001$ ). In the absence of significant linear contrasts, these intercepts indicate that the response to the probe was significantly larger than the response to the irrelevant stimuli. Because differential responding did not change over time, the mean of all 6 blocks was calculated and used

<sup>27</sup> Analysis of the unstandardized scores revealed a similar pattern. None of the measures (SCR, P300 amplitude, P300 recovery) showed a linear trend for any of the stimulus types (target, probe, irrelevant).

## CHAPTER 5



**Figure 5.2.** Grand average ERPs at Fz, Cz, Pz, and Oz as a function of stimulus type (target: recognized, admitted; probe: recognized, concealed; irrelevant: unrecognized; and null event: no stimulus presented).

in the subsequent analyses on detection efficiency and psychopathic traits. These averaged detection measures were uncorrelated with each other (see table 5.2).<sup>28</sup>

*Detection efficiency.* Signal detection parameters are presented in Table 5.1. The  $a$  and  $d$  values for z-SCR were .67 and 0.66, respectively, for z-P300 amplitude .68 and 0.50, respectively, and for z-P300 recovery .78 and 1.12, respectively. Combination measures of SCR and P300 were calculated by summing the  $z$ -scores for the individual measures. Combining SCR with P300 recovery resulted in an  $a$  and  $d$  value of .82 and 1.34, respectively, while the combination of SCR and P300 amplitude yielded an  $a$  and  $d$  of .73 and 0.76 respectively. For both P300 amplitude and P300 recovery, the increase in  $a$  caused by the addition of SCR was not significant. ( $z = 1.08$  and 0.88, respectively; see Hanley & McNeil, 1983)

<sup>28</sup> Analysis of the unstandardized scores also revealed no significant correlations between SCR and P300.



## COMBINING SCR AND SVT IN A CIT

**Table 5.1.** Means and standard deviations (SD) of the different detection measures and their combinations for the guilty participants and a simulated innocent group. Area under the receiver operating characteristic curve (*a*) with its 95% confidence interval. Standardized differences between the means of the two distributions (*d*).

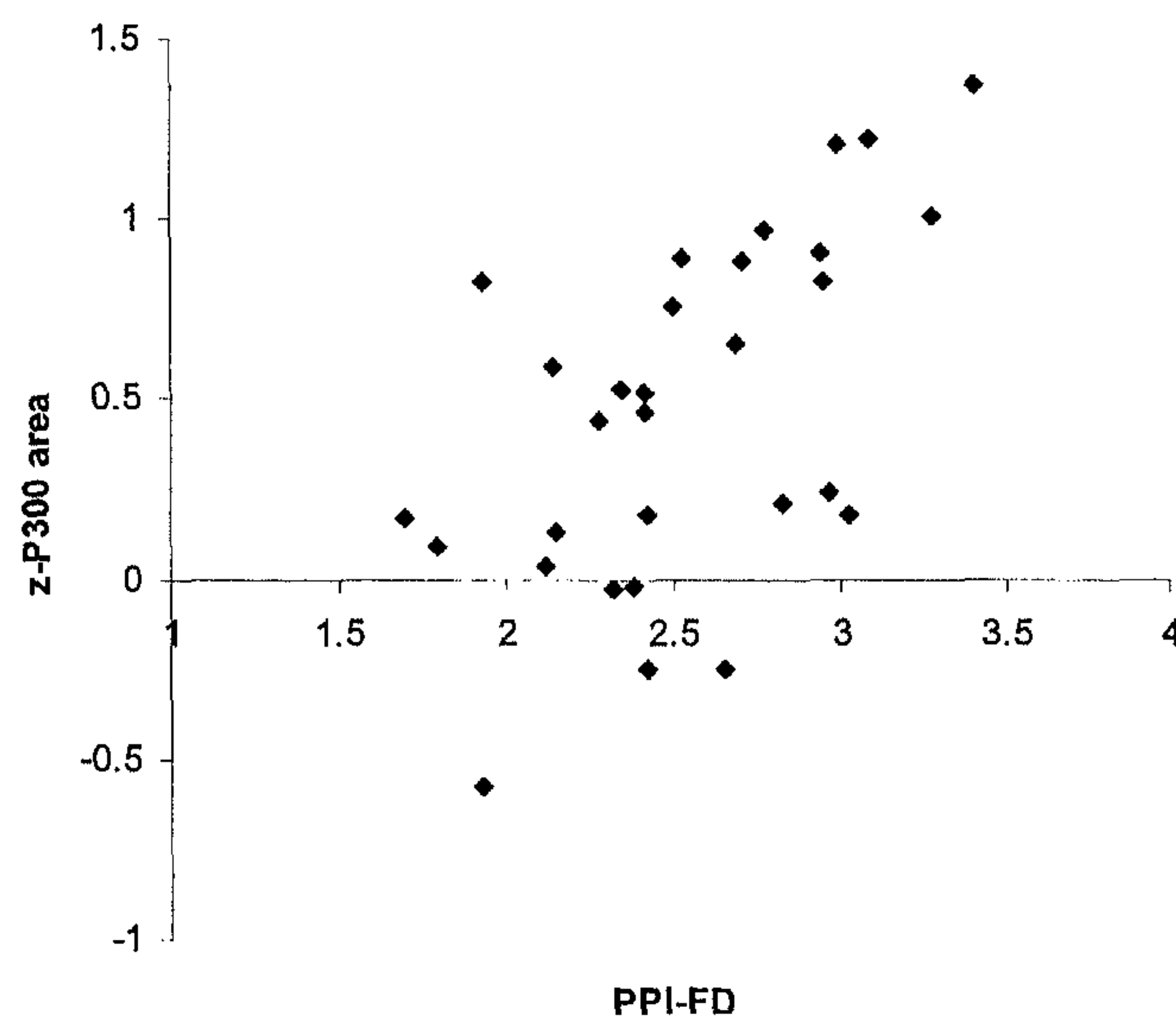
Detection measure	Mean z Guilty	SD Guilty	Mean z Innocent	SD Innocent	<i>d</i> '	<i>a</i>	95% CI of <i>a</i>
z-P300 amplitude	0.20	0.48	-0.02	0.41	0.50	.68	.54 - .82
z-P300 area	0.47	0.49	-0.02	0.39	1.12	.78	.66 - .90
z-SCR	0.21	0.45	-0.07	0.41	0.66	.67	.53 - .80
z- SCR + z-P300 amplitude	0.41	0.66	-0.09	0.66	0.76	.73	.60 - .86
z-SCR + z-P300 area	0.69	0.62	-0.09	0.54	1.34	.82	.72 - .93

**Table 5.2.** Correlations between the different detection measures and between the Fearless Dominance factor (PPI-FD) and the Impulsive Antisociality factor (PPI-IA) of the Psychopathic Personality Inventory (PPI) and the detection measures.

	z-P300 area	z-SCR	PPI-FD	PPI-IA
z-P300 amplitude	.02	.02	-.08	.03
z-P300 area	-	-.11	.58**	.12
z-SCR	-	-	-.06	.30
PPI-FD	-	-	-	.06

Notes. \*\**p* < .01

**Figure 5.3.** Scatterplot for the relationship between the Fearless Dominance subscale of the Psychopathic Personality Inventory (PPI-FD) and the standardized P300 area.



## CHAPTER 5

*Psychopathic traits.* Mean total PPI score was 364 ( $SD = 41$ ; range = 304 - 468). Internal consistency was high (Cronbach's  $\alpha = .88$ ). The Fearless Dominance and Impulsive Antisocial subscales were uncorrelated,  $r = .06$ ,  $p = .77$ . The z-P300 recovery measure correlated significantly with the Fearless Dominance subscale ( $r = .58$ ,  $p = .001$ ; see Figure 5.3). No other correlations reached significance (see Table 5.2).<sup>29</sup>

### 5.4 Discussion

The aim of this study was to determine whether psychopathic traits moderate SCR and P300 in a typical CIT. To this end, we simultaneously collected EEG and SCR while stimuli were presented in a higher order balanced sequence with a short ISI. The results indicate that the P300 recovery measure of concealed information correlated with psychopathic traits, whereas P300 amplitude did SCR did not.

The SCRs reported in this study were smaller than typical. Barry et al. (1993) also found that averaged SCRs collected in a paradigm with short ISIs are much smaller than SCRs collected with a long ISI. These authors also showed that, although smaller, these responses still exhibited the classic criteria of habituation, namely response decrement upon repeated presentation of the same stimulus (extinction), increased response to a change stimulus (response recovery), and to subsequent presentations of the original stimulus (dishabituation). Based on this, Barry and co-workers concluded that, even with short ISI paradigms, meaningful autonomic measures can be obtained. This is supported by data from the current experiment. The magnitude of the responses are similar to those found by Barry et al. (1993), and a significant difference between guilty and simulated innocent participants was found, indicating that our application of the m-sequences and subtraction of the null event were successful.

The positive correlation between psychopathic traits and z-P300 area replicates earlier work by Raine and Venables (1988), and is in line with the model proposed by Raine (1989), stating that psychopathic individuals have an enhanced ability to attend to stimulating events. Importantly, Raine (1989) speculated that this effect may be limited to stimuli that require active, effortful processing. In this light, it should be noted that a typical CIT resembles the continuous performance task used by Raine and Venables (1988) in that in both tasks the irrelevant stimulus type consists of

---

<sup>29</sup> No significant correlations between the unstandardized measures to any stimulus type and the PPI were found.



## COMBINING SCR AND SVT IN A CIT

multiple, physically different stimuli. In a typical oddball, however, all irrelevant stimuli are identical. This may explain why this link between psychopathy and P300 was not found in other studies, using typical oddball tasks (Kiehl et al., 2006; Kiehl et al., 1999). Meanwhile, Raine and Venables (1988) found that psychopathy was related to both P300 amplitude and P300 recovery time. In the current study, the correlation was limited to z-P300 area. Figure 5.2 illustrates that the increased z-P300 in the current experiment is, indeed, caused by a longer recovery time of the target.

The absence of a correlation between SCR and the P300 measures in the CIT indicates that these measures are independent of each other. This is in line with the results of Bahramali et al. (1997). Although these authors did find that target trials that elicited an orienting SCR showed a higher P300 than target trials on which such a response was absent, inspection of their data suggests that this effect was limited to the frontal and central midline site, and absent at the parietal midline site, which is the site we focused on in our study. That the coupling between orienting SCR and P300 is limited to the more frontal areas has also been shown in studies using single trial analysis of P300 and SCR (Woestenburg, Verbaten, & Slangen, 1981, 1983), and supports the notion that only the frontal, early P300 is an index of orienting (Soltani & Knight, 2000).

The  $d$  (0.66) and  $a$  (.67) values for the SCR measure were well below what is typically found in SCR based CIT studies (Ben Shakhar & Elaad, 2003). The specific task characteristics, data analyses, and the resulting small responses, however, make comparison with data collected in the traditional paradigm cumbersome, and the detection efficiency parameters should be interpreted with caution. More surprisingly, detection efficiency for the P300 measure was also lower than typically found. This can not be attributed to the task characteristics, which resembled a typical P300 based CIT. The most likely reason for this low detection efficiency is a lack of salience of the (probe) stimuli. Two recent studies (Rosenfeld et al., 2006; Rosenfeld et al., 2007) show that incidentally acquired knowledge, like details from a mock crime scenario, is harder to detect with P300 than more salient information like autobiographical stimuli (e.g., first and last names and birthdates), which is consistent with theoretical accounts of the P300 that incorporate stimulus meaning (Johnson, 1993).

Finally, combining the P300 with SCR did not significantly increase detection efficiency. The substantially lower detection efficiency found for the separate measures, however, makes it hard to draw firm conclusions related to this issue. Most importantly, the positive correlation between psychopathic traits and z-P300 area shows that P300 remains an important

## APPENDIX A

measure in the detection of deception, and may be a solution to the extent that psychopathic traits undermine autonomous responding in a CIT.

### Appendix A

Below is an example of a 3rd order balanced sequence:

<u>A A C</u>	<u>A B B C B A A C</u>	<u>A B B</u>
leading	Sequence	trailing
trials		trials

In this sequence, the 1<sup>st</sup> A is preceded by

A A C,

the 2<sup>nd</sup> A by

B C B,

and the 3<sup>rd</sup> A by

C B A

This means that at both the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> position preceding A each stimulus type appears once. The same holds true for all Bs in the sequence and consequently the average of A can be compared to the average of B



## Chapter 6: Combining Skin Conductance and Forced Choice in the Detection of Concealed Information

Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., & Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of deception. *Psychophysiology*, 44, 814-822.

### Abstract

An advantage of the concealed information polygraph test (CIT) is that its false positive rate is determined on statistical grounds, and can be set *a priori* at arbitrary low levels (i.e., few innocents declared guilty). This, however, inevitably leads to a loss of sensitivity (i.e., more guilty suspects declared innocent). We explored whether the sensitivity of a CIT procedure could be increased by adding an independent measure that is based on an entirely different psychological mechanism. In two experiments, we explored whether the accuracy of a CIT procedure could be increased by adding Symptom Validity Testing (SVT), a relatively simple, forced-choice, self-report procedure that has previously been used to detect malingering in various contexts. Results of a feigned amnesia experiment but not from a mock crime experiment, showed that a combination measure of both tests yielded better detection than either test alone.

## CHAPTER 6

### 6.1 Introduction

The use of the polygraph in criminal investigations has been heavily criticized in the scientific literature (e.g., Ben Shakhar, 2002; Fiedler et al., 2002; Lykken, 1998; National Research Council, 2003). This critique primarily concerns the Control Question Test (CQT), the technique most widely used in police investigations. During a CQT, physiological recordings to questions directly related to the incident under investigation (e.g., "Did you stab John Doe?") are compared to emotionally provocative control questions (e.g., "During the first 20 years of your life, did you ever hurt someone physically?"). Stronger physiological responding to the control questions is taken as an indication of innocence, whereas stronger physiological responding to the questions directly related to the crime is taken as an indication of deception. The CQT has been criticized for its lack of theoretical framework, its lack of standardization, and the fact that it relies on improper controls, resulting in a high percentage of false positives (i.e., innocent examinees tested guilty; see Lykken, 1998; Ben-Shakhar, 2002).

A different interrogation technique, first described by Münsterberg (1908), and later named the Guilty Knowledge Test (GKT; Lykken, 1959, 1960) or Concealed Information Test (CIT) is used less frequently. In fact, the only country where it is employed on a large scale is Japan (Hira & Furumitsu, 2002; Nakayama, 2002). In a CIT, physiological measures are similar to the CQT, but questions are presented in a multiple choice format (e.g., "Was the amount of money stolen \$10.000?, \$20.000?, \$30.000?, \$40.000? or \$50.000?"). All questions concern intimate details of the crime, of which only the investigative authorities and the perpetrator are presumed knowledgeable. Consistent stronger physiological responding to the correct answers reflects knowledge of these details and thus involvement in the crime.

Unlike the CQT, the CIT is highly standardized and has a sound theoretical underpinning in orienting theory, as recently shown by Verschuere, Crombez, De Clercq, and Koster (2004). It has also been shown to be a robust tool for discriminating between guilty and innocent participants. A recent meta-analysis on 80 studies revealed an average effect size ( $d$ ) of 1.55, and this effect size was even larger (3.12) under optimal conditions (i.e., motivational instructions, deceptive verbal response, at least 5 questions; Ben Shakhar & Elaad, 2003). Furthermore, the CIT provides adequate safeguards for the innocent in that the probability of a false positive outcome can be determined *a priori*. This probability depends on



## COMBINING SCR AND SVT IN A CIT

the number of questions, the number of alternative answers per question, and the criterion for guilt. For example, when a CIT contains 4 questions with 4 answers each, the probability for an innocent examinee to systematically show the strongest physiological response to the correct alternative is  $(1/4)^4 = .004$ . If the criterion for guilt is set at 'respond maximally to at least three out of the four questions', this probability becomes .05.<sup>30</sup> Thus, for a CIT, the probability of a false positive outcome can be set *a priori* by including sufficient questions and alternatives, and by selecting a proper detection criterion. For use in criminal justice, detection criteria that result in a low number of false positive outcomes (i.e., high specificity) are important, since it adheres to the legal doctrine in most countries, abbreviated in the so-called Blackstone Maxim: "Better that ten guilty persons escape than that one innocent suffer" (Blackstone, 1882; Volokh, 1997).

However, an inevitable consequence of setting the detection criteria at levels corresponding with high specificity is that the percentage of false negative outcomes (i.e., a guilty examinee tested innocent) becomes larger. Several recent studies have estimated this percentage to range between 14% and 24% (Ben Shakhar & Elaad, 2003; Elaad, 1998). These studies primarily relied on mock crime scenarios to determine detection efficiency. The false negative rates in field studies on detection efficiency of the CIT are even higher. For example, Elaad (1990) and Elaad, Ginton, and Jungman (1992) found that the CIT missed approximately half of the guilty suspects, when using skin resistance response as the detection measure and confessions as an index of guilt. A more recent mock crime study by Carmel, Dayan, Naveh, Raveh, and Ben-Shakhar (2003) showed that under realistic conditions, the percentage of false negative outcomes was as high as 48%. The problem of false negative outcomes associated with the CIT cannot simply be explained by a perpetrator's failure to remember pertinent details. Elaad (1990), for example, presented the items from a selection of the records to independent judges. The judges were asked whether these items had at least an 80% likelihood of being recognized by the guilty subject. This was the case in the majority of the items, leading Elaad to conclude that the high proportion of false negatives was due to the low number of questions, rather than items not having been encoded in memory. Likewise, in the study by Carmel et al. (2003), even though the false negative rate was

---

<sup>30</sup> This chance is composed of the non-mutually exclusive events of responding maximally to all 4 questions (A) and responding maximally to three out of the four questions (B), and can be calculated as follows:  $P(A) + P(B) - P(A \text{ and } B)$ , and equals  $.25^4 + 4 \cdot .25^3 - 4 \cdot .25^4 = .0508$ .

## CHAPTER 6

only 10% under optimal conditions (i.e., a CIT containing only questions on central details, performed immediately following a realistic mock crime), this rate still ranged from 20 to 45% when only correctly recalled items were analyzed.

One approach to reduce the number of false negatives without sacrificing specificity is combining several highly specific measures. So far, this approach has primarily focused on the addition of new psychophysiological indices to the standard measurement of skin conductance. For example, a specific respiration parameter (respiration line length; RLL) has been shown to increase the sensitivity (i.e., proportion of correctly classified guilty examinees) of a skin conductance based CIT (e.g., Ben-Shakhar & Dolev, 1996; Ben-Shakhar & Elaad, 2002; Ben Shakhar, Gronau, & Elaad, 1999; Elaad et al., 1992; Timm, 1982). More recently, measures such as Finger Pulse Waveform Length (Elaad & Ben-Shakhar, 2006) and Normalized Pulse Volume (Hirota, Sawada, Tanaka et al., 2003) have been proposed as useful additional CIT parameters. In general, the addition of extra physiological variables is useful if the false negative outcomes of the CIT are due to noise in the measurement of the underlying psychophysiological mechanism (i.e., the orienting response). In that case, increasing the number of physiological indices that are manifestations of this underlying mechanism will reduce the noise and hence the percentage of false negatives.

If the false negative outcomes are not due to measurement noise, but simply result from the absence of response in the underlying mechanism, adding more psychophysiological indices might be less fruitful. In this case, simply adding more physiological indices will not increase sensitivity. It might then be more fruitful to combine the CIT with highly specific measures that are independent of the psychophysiological mechanism tapped by CIT (see also Nies & Sweet, 1994). An interesting candidate measure that might be used in this way is the Symptom Validity Test (SVT). This test was developed to detect malingering and has been used in a wide variety of fields, including detection of malingering of perceptual deficits (Brady & Lind, 1961), short-term memory deficits (Hiscock & Hiscock, 1989), amnesia for specific events (Frederick, Carter, & Powel, 1995), and cognitive deficits attributed to chronic pain (Meyers & Diep, 2000) or posttraumatic stress disorder (Rosen & Powel, 2003).

The rationale of the SVT lies in the notion that the performance of honest individuals (i.e., individuals with genuine perceptual or memory deficits or without intimate knowledge of a crime) on a forced choice test will be at chance level. Consider, for example, an individual with genuine



## COMBINING SCR AND SVT IN A CIT

color blindness. This person is presented with a series of events, for example red or blue illuminating bulbs. After each event, the individual is asked to name the color of the light bulb that lit up. These questions have a forced-choice format with answers of equal probability (e.g., 1.) blue, 2.) red). For this person, test performance will be at chance level. An individual with intact perception will perform above chance level. Deception (i.e., malingering of a perceptual deficit) is inferred when performance falls significantly below chance level as the person apparently has the ability to systematically avoid the correct answers and select the incorrect answer more often than predicted by chance (Denney, 1996). Even when aware of this rationale, dishonest people may still fail the test due to humans' incapacity to generate random series of responses (see also Haughton, Lewsley, Wilson, & Williams, 1979; Wagenaar, 1972). Because the distribution of the number of correct answers in individuals with no true ability is known (binomial), the SVT, like the CIT, allows for the computation of the probability of a false positive outcome at any chosen detection criterion.

Denney (1996) adapted the SVT for use in a forensic setting (see also Lieblich & Ninio, 1972; Lieblich, Shaham, & Ninio, 1976). He described three cases where the defendant claimed amnesia for their crime. In all cases, defendants performed well below chance level at an SVT consisting of questions concerning intimate details of the crime (e.g., "How did the perpetrator leave the bank? 1.) walking, 2.) running"). More recent research on the accuracy of the SVT in detecting feigned amnesia for mock crimes shows that, at specificity levels of 95 %, its sensitivity ranges from 40 to 60% (Jelicic et al., 2004a, 2004b; Merckelbach et al., 2002).

Because the SVT measures a different psychological mechanism than the CIT (i.e., limitations of the cognitive system in producing random-like responses versus an orienting response), combining it with a CIT may decrease the number of false negative outcomes, while maintaining high levels of specificity. The aim of the current experiments was twofold: first, we explored whether the SVT can be used as an indicator of deception. Second, we investigated whether the addition of SVT has an incremental value beyond that of a skin conductance based CIT. In the first experiment, participants performed a mock crime, mimicking the application of the CIT and SVT in a typical forensic application. In the second experiment, participants were instructed to feign complete amnesia of their identity, after which they were given a CIT and SVT containing biographical data. This test mimics deception in a somewhat different context, for example to

## CHAPTER 6

judge the veracity of amnesia claims, or when someone's identity is the topic of investigation.

### *6.2 Experiment 1*

#### 6.2.1 Method

##### *Participants*

Participants were 65 undergraduates who received either course credits or a small financial compensation. Five participants were excluded from data analyses because they failed to follow instructions, or because of equipment malfunctioned. Thus, the remaining sample consisted of 60 students (20 men) with a mean age of 21 years ( $SD = 2.8$ ). All participants read and signed a letter of informed consent before participating. The experiment was approved by the Faculty's ethical committee.

##### *Physiological Measures*

Skin conductance was measured using a 24 bit DC 0.5 Volt system. Two Beckmann silver/silver chloride (Ag/AgCl) electrodes (5 mm in diameter) were placed on the medial phalanges of the first and second fingers of the participants' non-dominant hand. Electrodes were filled with isotonic electrode paste (0.9% NaCl). Respiration was measured using a strain gauge attached around the thorax. All data were acquired using Contact Precision Instruments bioamplifiers with a sample rate of 200 Hz.

##### *Procedure*

Upon arrival in the laboratory, the participant was given written instructions to carry out a scenario. For half of the participants ( $N = 30$ ), these instructions entailed the guilty scenario. The other half received the innocent scenario. The guilty scenario consisted of stealing 20 euros and a mobile telephone hidden away in a jacket in a café located inside the university building. In order to gain access to this café, participants had to collect a key, which was located in a drawer of a kitchen unit. In the innocent condition, the task involved collecting a dirty cup from a kitchen and washing it elsewhere. Both innocent and guilty instructions were concluded by telling the participant to wait for further instructions in a waiting room. After 15 min the examiner entered the room and informed participants that "a crime has been committed and you are one of the suspects. If you are guilty, try to lie effectively during the lie detection test so that you will be declared innocent." Following this, the experimenter escorted the participants back to the laboratory where testing commenced.

The CIT consisted of one example question and 6 genuine questions. Questions were presented on a 15 inch monitor. Each question was followed by a set of 6 items, among which was the correct answer



## COMBINING SCR AND SVT IN A CIT

(critical item). The first item was never the critical item and served to absorb novelty orienting responses. The 6 questions of the CIT addressed both central and peripheral details of the crime. Each question was displayed for 10 s. Then, a blank screen followed for 3 s, after which the first item was displayed for 3 s. Next, another blank screen followed for 10 s. This cycle was repeated for each of the 6 items, creating a 26 s inter stimulus interval. The critical item was always positioned at either the 3rd, 4th, or 5th place. The order of the questions was determined by a balanced latin square. Participants had to respond to the presentation of each item with a verbal “no” answer. A participant-terminated break was given after completion of an entire question.

Upon completion of the CIT, participants were given the SVT. The SVT consisted of 12 questions, each with two equally plausible alternatives. These items were checked using a Doob and Kirschenbaum pilot procedure to ensure they were all equally plausible (Doob & Kirschenbaum, 1973). For this procedure, 10 naïve participants were given all questions, and asked to pick the most plausible item. Any item for which the probability was below .3 or above .7 was discarded. Six of the 12 SVT questions resembled those of the CIT. The other 6 questions concerned specific details of the café where the mock crime took place. The SVT was administered in the form of a booklet, containing only one question per page with the following instructions: “Complete this questionnaire by circling one of the answers to each question. You must always choose one option. If you do not know the answer, just guess. You must answer the questions in the order they are presented. Do not turn to the next page unless the question has been answered. Do not turn back the page under any circumstance.” The thickness of the booklet was increased by adding twelve empty pages at the end. This was done so as to obscure the true length of the test, making it difficult for participants to calibrate their performance in accordance with chance. To prevent participants from deriving correct CIT answers from the SVT answers, the CIT was always administered first.

All testing took place in a dimly lit, sound-proof, air-conditioned laboratory. Participants were monitored from a control room by means of a video surveillance camera and a microphone.

### *Response scoring and data analysis*

The maximal positive deflection in skin conductance during the 1 s to 5 s interval after stimulus onset was defined as the SCR. To eliminate individual differences in responsivity, within question standardized scores were computed by subtracting the mean of all 5 responses from the response to the critical item and dividing that by the standard deviation of all 5

## CHAPTER 6

responses (Ben-Shakhar, 1985). These standardized scores were then averaged over questions in order to produce a single detection score for the CIT.

Siegel's (1956) formula was used to calculate the z-score for the SVT:  $z = ((x \pm 0.5) - NP) / \sqrt{NP(1 - P)}$ . Here,  $z$  is the test statistic;  $x$  is the number of correct responses;  $N$  is the total number of questions (i.e., 12);  $P$  is the probability of a correct discrimination given no true ability (i.e., 0.5). Due to the fact that the binomial distribution involves a discrete variable, a correction for continuity was made: adding 0.5 when  $x < NP$ , and subtracting 0.5 when  $x > NP$ .

### 6.2.2 Results and discussion

The SVT and CIT scores within guilty participants were uncorrelated ( $r = -.11$ ;  $p = .56$ ). In order to derive accuracy rates, cutoff points for the detection measures were set at a  $z$ -score  $< -1.65$  for the SVT (corresponding to a specificity of 95%), whereas for the CIT, the Lykken score was used. With the latter, each question is assigned 2 points if the response to the crime relevant item is the largest of all responses, 1 point is assigned if it is the one but largest, while 0 points are assigned in all other cases. All points are then added, and a score of 6 or more is taken as a guilty test outcome. Based on the binomial theorem, this cutoff point corresponds to a specificity of 83% (see also MacLaren, 2001). In order to determine accuracy rates for the combination of the CIT and SVT, we used the Independent Parallel Testing approach (National Research Council, 2003 p. 367). With this approach, deception is inferred if any of the individual tests is positive. Consequently, overall test outcome is negative only when all individual tests are negative.

The cutoff resulted in correct classification of all of the innocent (100%) and 14 (47%) of the guilty participants for the CIT. For the SVT, it yielded correct classification of all (100%) of the innocent and 8 (27%) of the guilty participants. The combination of the CIT and SVT resulted in correct classification of all innocent (100%) and 17 (57%) of the guilty participants.

Defining guilt and innocence using the criteria based method described above has the disadvantage that it relies on a single, arbitrary cutoff point. An alternative approach to describing detection efficiency that does not have this disadvantage is signal detection theory (SDT; National Research Council, 2003). This method defines detection efficiency in terms of the degree of separation between the distributions of the detection measure for the innocent and the guilty condition. In order to produce a



## COMBINING SCR AND SVT IN A CIT

**Table 6.1.** Means and standard deviations (SD) of the Concealed Information Test (CIT), Symptom Validity Test (SVT), and their combination for the guilty and innocent condition. Standardized differences between the means of the guilty and innocent condition ( $d'$ ). Area under the receiver operating characteristic curve ( $a$ ), with its 95 % confidence interval.

Measure	Mean z Guilty	SD Guilty	Mean z Innocent	SD Innocent	$d'$	$a$	95 % CI of $a$
Experiment 1							
CIT	.78	.61	-.06	.41	1.62	.86	.77 - .96
SVT	.73	1.41	-.15	.71	.79	.70	.56 - .83
SVT & CIT	1.51	1.47	-.21	.88	1.42	.84	.74 - .95
Experiment 2							
CIT	.95	.60	.10	.39	1.68	.88	.81 - .94
SVT	2.08	1.34	.02	.76	1.89	.87	.81 - .93
SVT & CIT	3.03	1.44	.12	.85	2.46	.95	.91 - .99

single detection score for the combination of the two tests, the SVT  $z$ -score was multiplied with -1 and added to the CIT  $z$ -score. Subsequently, the distance between the centers of the distribution of the innocent and the distribution of the guilty was computed in terms of standard deviation ( $d'$ ), and the area under the (empirical) Receiver Operating Characteristic (ROC) curve ( $a$ ) was computed. These statistics are presented in Table 6.1 (top panel). This table shows that  $d'$  values for the CIT and SVT were 1.62 and 0.79, respectively. The  $d'$  value for the combination of CIT and SVT was 1.42. The areas under the ROC curve were .86 for the CIT, .70 for the SVT, and .84 for their combination.

These results indicate that the SVT can be used to detect deception in a typical forensic setting, even though its sensitivity is modest. The signal detection parameters revealed no incremental validity of the SVT over the CIT. In order to conceptually replicate these results, we conducted a second experiment.

In this second experiment, a number of methodological improvements and extensions were made. First, to allow for generalization of the results, a community sample was used, an incentive for beating the test was given, and a feigned amnesia paradigm was applied. To increase statistical power, only guilty participants were included. Furthermore, any possible carryover effect due to the fixed order in Experiment 1 was addressed by using different questions for each test, and balancing the order of the tests. Also, the Psychopathic Personality Inventory (PPI; Lilienfeld & Andrews, 1996) was included as a measure of psychopathic traits.

Previous research has shown that psychopathy might be a moderating factor in detecting concealed information. Hyporeactivity is a prominent feature of psychopathy (Lorber, 2004), and recently Verschuere

## CHAPTER 6

et al. (2005) showed that prisoners who score high on certain subscales of the PPI exhibit both a decreased overall electrodermal orienting response, as well as decreased differential electrodermal orienting responses to the correct and incorrect CIT items. The PPI was included to examine whether this hyporeactivity phenomenon in high PPI individuals could be replicated, and if addition of a SVT could be a potential solution for the reduced detection efficiency that it implies.

### ***6.3 Experiment 2***

#### ***6.3.1 Method***

##### *Participants*

Participants were 60 persons (18 men) recruited through advertisement in local newspapers. The mean age was 33 years ( $SD = 9.1$ ). All participants read and signed a letter of informed consent before participating. The experiment was approved by the Faculty's ethical committee.

##### *Measurements*

The Dutch translation of the Psychopathic Personality Inventory (PPI; Jelicic et al., 2004; Lilienfeld & Andrews, 1996) was used to assess psychopathic traits. Following Benning, Patrick, Hicks, Blonigen, and Krueger (2003) and Verschuere et al. (2005), we calculated the Fearless Dominance factor and the Impulsive Antisocial factor by summing scores across the appropriate subscales while compensating for the fact that these subscales consist of different numbers of items. Physiological measures were identical to those in Experiment 1.

##### *Procedure*

Participants who responded to the newspaper advertisement were contacted for an appointment. During this initial contact, they were asked to supply autobiographical information (e.g., place of birth, mothers maiden name etc.) that was subsequently used as stimulus material in the experiment.

Upon arrival in the laboratory, participants were asked to fill out the PPI. Subsequently, they were given written instructions. These instructions explained that in some circumstances, claiming memory problems can have beneficial effects. An example of how feigning memory problems after a traffic accident could increase compensation payments paid by the insurance company was given. Next, participants were instructed to feign complete memory loss of their identity and told that the experiment was designed to test new methods to detect their deceit. They were explicitly told to try to beat the test and were promised a €5 reward if they succeeded.

Initially, participants provided the experimenter with a possible total of 24 autobiographical details. On the basis of these details, 18 questions



## COMBINING SCR AND SVT IN A CIT

were constructed, divided into 3 sets of 6 questions, such that the different sets all contained questions of similar nature (e.g., each set contained the same number of names of relatives). Subsequently, 1 set was used for the CIT and 2 for the SVT. The order of the two tests was counterbalanced. The remainder of the procedure was identical to Experiment 1, with the exception that the critical alternative was randomly presented at any position except for the first.

### *Response scoring and data analysis*

Response scoring and data analysis were similar to Experiment 1. Since only 'guilty' participants were included, signal detection parameters were derived differently (see below).

### 6.3.2 Results and discussion

SVT and CIT scores within guilty participants were uncorrelated ( $r = -.04$ ,  $p = .77$ ). Using identical criteria as in Experiment 1 resulted in correct classification of 39 (65%) of the participants for the CIT and 38 (63%) of the participants for the SVT. The combination of the CIT and SVT correctly classified 53 (88%) of the guilty participants.

In order to derive signal detection indices, a number of previous studies that included only guilty participants compared the distribution of the standardized critical items to the distribution of the average standardized control items (e.g., Elaad & Ben-Shakhar, 2006; Gronau et al., 2005; Verschuere et al., 2005; see also Ben-Shakhar, 1985). As we will argue below, however, this procedure is suboptimal, and overestimates detection efficiency.

The standardization procedure described by Ben-Shakhar (1985) entails subtracting the mean and dividing the outcome by the standard deviation of responses to all alternatives from either the response to the critical alternative or from the response to the control alternatives. As a consequence, all information in the dataset is used to derive the distribution of the critical item. Applying the same procedure to the control items can thus, by definition, not result in a distribution of control items containing unique information. In fact, each participant's score for the control items, together forming the 'innocent' group, is linearly dependent on that participant's score on the critical items. This is because the score on the control items is simply the score on the critical item, divided by  $-(N-1)$ , where  $N$  denotes the total number of unique stimuli. For the demonstration please refer to appendix A. Furthermore, as a consequence of the averaging over the standardized control items, the standard deviation of the distribution of control-scores becomes approximately  $\sqrt{N-1}$  times smaller than would have been the case if they had been obtained from a single item

## CHAPTER 6

(e.g., the item that was critical in the 'guilty' group). Thus, in case of successful detection, this procedure renders a distribution of the control item with a negative mean and a smaller standard deviation than a hypothetical group of truly innocent participants. The latter would have a mean of 0 and an unbiased standard deviation. As a consequence of this negative mean and smaller standard deviation, signal detection parameters will be unjustly inflated.

Alternatively, we chose to base our signal detection parameters on a comparison with a simulated innocent group consisting of 60 participants (see also Carmel et al., 2003). Such a group was created for both tests by randomly drawing values from their respective distributions, and treating these values in exactly the same manner as the values measured for guilty participants. For the SVT, this entailed drawing 60 values from the binomial distribution with  $N = 12$  and  $P = .5$ . For the CIT it entailed the following steps. First, five values were randomly drawn from a standard normal distribution (mean = 0, standard deviation = 1). Then, one value (as the 'response' to the critical item) was standardized relative to the mean and standard deviation of all five responses. This way, a standardized score for one innocent person for one question was derived. This process was repeated six times (to simulate six questions), and these six values were averaged to represent a score for one innocent participant. Based on this procedure, the  $d'$  values were 1.68, 1.89, and 2.46 for the CIT, SVT and their combination respectively, while  $a$  values were .88, .87, and .95. These values, with their corresponding 95% confidence interval are presented in Table 6.1 (bottom panel). Statistical testing of these  $a$  values revealed that the combination of the two tests outperformed the CIT alone ( $z = 2.24$ ,  $p = .01$ ; see Hanley & McNeil, 1983).

Fifty-seven participants filled out the PPI completely. Mean total score was 341 ( $SD = 39$ ; range = 256-420). Internal consistency was high (Cronbach's  $\alpha = .89$ ). The Fearless Dominance and Impulsive Antisocial subscales were uncorrelated,  $r = .17$ ,  $p = .21$ . To investigate the relationship between psychopathic personality traits and overall physiological responding, we computed the correlation between these PPI subscales and the unstandardized mean skin conductance response. Neither the Fearless Dominance factor nor the Impulsive Antisocial factor significantly correlated with overall physiological responding ( $r = -.08$ ,  $p = .57$  and  $r = -.04$ ,  $p = .75$ , respectively). Similarly, the correlations between the two subscales and the detection measures for CIT and SVT did not attain significance ( $r = -.24$ ,  $p = .07$  and  $r = -.11$ ,  $p = .41$  for CIT and  $r = -.07$ ,



$p = .62$  and  $r = -.11$ ,  $p = .42$  for SVT with the Fearless Dominance and the Impulsive Antisocial subscale, respectively).

#### **6.4 General discussion**

This study aimed to investigate whether SVT can be used to detect deception and whether combining it with a CIT would yield detection efficiency superior to that of the CIT alone. First of all, the results from both experiments show that the SVT can be used to detect deception. Furthermore, we found that combining the two tests yielded superior detection efficiency, but only in the feigned amnesia experiment.

In both experiments, the accuracy rate of the SVT in detecting deception was similar to that found in studies on false claims of amnesia (e.g., Jelicic et al., 2004a; Jelicic et al., 2004b; Merckelbach et al., 2002). This is not surprising since the instructions to the participants in all these studies were highly similar. That is, the instruction to feign amnesia for a mock crime is a special instance of instructing participants to lie (Christianson & Merckelbach, 2004).

The accuracy rates obtained with the CIT were equivalent to those found in earlier studies as well (e.g., Carmel et al., 2003). Importantly, the instructions to the guilty participants in our Experiment 1 contained no specific information addressed by the subsequent CIT items (e.g., the instructions read ‘steal the money...’ and not ‘steal the 20 euro...’). Therefore, our Experiment 1 would qualify as a procedure that Carmel and coworkers (2003) termed a ‘valid mock crime’, and it yielded similar detection rates as obtained by these authors.

The ability of both the SVT and CIT to differentiate between guilty and innocent participants is also evident from the  $d'$  values. In terms of Cohen (1988), the value of 0.79 for the SVT in Experiment 1 represents a moderate to large effect size, whereas the 1.62 for the CIT in Experiment 1, and the values of 1.68 and 1.89, respectively found in Experiment 2 all represent a large effect size.

The detection efficiency found in Experiment 2 was higher than that in Experiment 1, particularly for the SVT. This may explain why the incremental validity was limited to Experiment 2. The predictive validity of the SVT in Experiment 1 may simply have been too low to establish a significant incremental validity given the number of participants.

There are three factors that may have contributed to the difference in detection efficiency between the two experiments. To begin with, it might be that because of its personal relevance, the autobiographical paradigm yields higher accuracy than the mock crime paradigm. At first

sight, this might seem difficult to reconcile with the findings of the Ben-Shakhar and Elaad (2003), who found that mock crime studies yield higher accuracy than studies using the personal item paradigm. It should be noted, however, that our personal item paradigm was adapted such that participants were specifically instructed to feign amnesia whereas in many other studies, participants are merely instructed to deny recognition. This way our personal item paradigm more closely resembles a mock crime than a typical personal item paradigm. Furthermore, in their study on the validity of reaction times in the detection of concealed information, Gronau et al. (2005) found a similar pattern. In that study, reaction times differed between relevant and irrelevant items when they denoted personally significant information, but not when they pertained to mock crime details. Second, participants in Experiment 2 were promised a financial incentive for beating the test. This may have increased accuracy through an increase in motivation, and is probably ecologically valid, since in typical applied settings, great interests are at stake. Finally, the fact that participants in Experiment 2 were drawn from a community sample may have boosted detection rates. After all, people from a community sample are less likely to understand the rationale of the SVT and they have less knowledge of the phenomenon of random performance. However, both Jellicic et al. (2004b), and Merckelbach et al. (2002) also found sensitivities in the order of 60% in an undergraduate sample.

The continuous scores on the CIT and SVT within the guilty group were independent ( $r = -.11$  in Experiment 1 and  $r = -.04$  in Experiment 2). Also the dichotomized scores ('innocent' vs 'guilty') showed independence: it appeared that the actual probability of being declared guilty on either test (.57 in Experiment 1, .88 in Experiment 2) was very close to the expected probability given independence of the probability of a guilty SVT outcome and the probability of a guilty CIT outcome (.61 in Experiment 1, .87 in Experiment 2).<sup>31</sup> One implication of this independence is that, to the extent that false negative outcomes of the CIT are caused by too small responses in the underlying psychophysiological mechanisms, adding SVT might be a solution. A second practical implication is that, since the two tests measure different mechanisms, they may not be susceptible to the same countermeasures. Countermeasures refer to anything that an individual might do in an effort to defeat or distort a polygraph test (Honts, Devitt,

---

<sup>31</sup> If the probabilities of being declared guilty on each test are independent of each other, under the parallel testing rule, the expected probability of being declared guilty on either or both is given by:  $seA + (1 - seA) \times seB$ , where  $seA$  represents the sensitivity of test A and  $seB$  represents the sensitivity of test B.



## COMBINING SCR AND SVT IN A CIT

Winbush, & Kircher, 1996). To the extent that various physiological indices are manifestations of the same underlying mechanism (i.e., orienting response), any countermeasure aimed at interfering with this mechanism (e.g., counting backwards from 200 by 7) is likely to have similar undermining effects on these indices. In contrast, adding independent tests may limit the effects of such countermeasures, although this is an issue that warrants systematic empirical testing.

The psychopathy scores obtained in Experiment 2 did not show any link to overall psychophysiological reactivity, nor with the detection measure of the CIT. This failure to replicate the findings of Verschuere et al. (2005) may have various reasons. For one thing, Verschuere et al. (2005) had a prison sample, whereas the current study relied on a community sample. Although one would expect more extreme psychopathy scores in a prison sample, inspection of the data does not confirm this. The mean total score of the community sample in Experiment 2 ( $M = 341$ ,  $SD = 39$ ) was not dramatically lower than that reported by Verschuere and coworkers for their prison sample ( $M = 350$ ,  $SD = 40$ ). Another possible explanation might be that in our second experiment, contrary to the study by Verschuere et al. (2005), a monetary incentive was promised. It could be argued that psychopaths do not show underarousal under motivational conditions (Arnett, 1997; Verschuere et al., 2007). Furthermore, our failure to find an association between psychopathy scores and psychophysiological reactivity is in line with earlier work of Gudjonsson (1982) and Balloun and Holmes (1979), who also found no effect of personality on the detection of concealed information. The exact relation between the PPI and autonomous reactivity remains unclear, and merits further research.

Another important point of consideration is the practical applicability of the SVT. When combining this procedure with a CIT, little extra effort is needed. Assuming that in the process of preparing CIT questions, the crime scene has been visited or the criminal records and files have been inspected, little extra effort is needed to create additional SVT items. Some authors (e.g., Podlesney, 1993, 2003) have argued that in real-life cases, it is often difficult to formulate sufficient questions with equally plausible answer items. On the other hand, SVT items can be constructed using only 2 plausible answer alternatives. They allow, for example, yes/no answer options. This makes it relatively easy to develop sufficient items.

When using these tests as forensic tools, one needs to keep in mind that with the cutoff points chosen, and even with SVT and CIT combined, specificity is higher than sensitivity. This implies that both measures can best be used as challenge tasks in the forensic domain. Thus, failing to pass

## CHAPTER 6

the SVT or the CIT is a strong indication of guilt, but passing both tests is not a strong indication of innocence (see also Denney, 1996).

In the analysis of the data from Experiment 2 we encountered a problem with the method that is often used to derive signal detection parameters. As a solution, we proposed a simulated 'innocent' group. Comparing the two methods by cross checking our own data from Experiment 1 supported our conclusion that an analysis based on the guilty participants' standardized critical items and the guilty participants' standardized control items leads to an overestimation of the signal detection parameters. This method yielded an  $a$  of .95 and a  $d'$  of 2.21, while the empirical  $a$  and  $d'$  were .86 and 1.62 respectively. Comparison of the standardized critical item of the guilty participants to a simulated group of innocent participants yielded an  $a$  of .85 and a  $d'$  of 1.55, indicating that our simulation procedure yields better estimates of the fully empirical signal detection parameters. It is also noteworthy that the mean and standard deviation of the empirical group of innocents in Experiment 1 and the simulated group of innocents in Experiment 2 were highly similar. The overestimation of signal detection parameters that results from Ben-Shakhar's (1985) method might also explain the high values found in the study by Gronau et al. (2005), and might also explain why Elaad and Ben-Shakhar (2006) found better detection efficiency in their experiment with only guilty participants than in their experiment including guilty and innocent participants. We recommend that future research using Ben-Shakhar's (1985) standardization procedure should either incorporate both guilty and innocent subjects, or compare the distribution of the guilty participants to that of a simulated group of innocent participants.

Finally, we can make two recommendations for future research. First, data from Experiment 2 showed that addition of a behavioral measure increased detection efficiency. Addition of a respiration measure, however, has also been shown to increase detection efficiency (Ben-Shakhar & Dolev, 1996; Elaad et al., 1992; but see Verschuere et al., 2007). Future studies could make a direct comparison between the incremental validity of psychophysiological measures like respiration with that of behavioral measures like the SVT. Secondly, recent studies have shown that peripheral details do not serve as good CIT items (Carmel et al., 2003; Jokinen, Santtila, Ravaja, & Puttonen, 2006). Of special interest in this context is a study by Lieblich, Ben-Shakhar, and Kugelmass, (1976), who showed that detection of relevant autobiographical information like names of relatives was better than detection of less relevant information like favorite brand of cigarettes. To the extent that the detection efficiency of the CIT is determined by



## COMBINING SCR AND SVT IN A CIT

salience of the test items, reserving the most salient items for the CIT, while including the less salient items in the SVT may even boost detection efficiency more than found in Experiment 2.

Altogether, our results suggest that it is worthwhile to combine several different types of lie detection, and that testing for concealed information needs not be confined to indices measuring orienting response.

## CHAPTER 6

### Appendix A

The purpose of this appendix is to demonstrate the linear dependency of the standardized response to a critical item (probe) and the standardized response to control items (irrelevants), when both are derived from the same data set.

For each participant and question, we measured the responses to  $N = 5$  unique stimuli, that is 1 probe ( $p$ ), and  $N-1$  irrelevants ( $i_1, \dots, i_{N-1}$ ). Following Ben-Shakhar (1985), these responses were transformed to  $z$ -scores for the irrelevants:

$$z_j = (i_j - X) / \text{sd}_x, \quad j = 1, \dots, N - 1, \quad (1)$$

and for the probe:

$$z_p = (p - X) / \text{sd}_x, \quad (2)$$

where  $X$  and  $\text{sd}_x$  denote the average and standard deviation across all  $N$  stimuli, respectively.

In order to derive an ROC curve, and compute the area under this curve ( $a$ ) and  $d'$ , Ben-Shakhar used the responses of 'guilty' subjects to the irrelevant answer-stimuli to generate an 'innocent' group (representing a group for whom all answer-stimuli would be irrelevant). For this, the average of the irrelevant  $z$ -scores was used:

$$Z_i = \sum (z_1, \dots, z_{N-1}) / (N - 1). \quad (3)$$

By definition, the average of all  $z$ -scores is 0:

$$\sum (z_1, \dots, z_{N-1}, z_p) / N = 0. \quad (4)$$

This means that:

$$\sum (z_1, \dots, z_{N-1}, z_p) = 0, \quad (5)$$

and also that:

$$\sum (z_1, \dots, z_{N-1}) + z_p = 0. \quad (6)$$



## COMBINING SCR AND SVT IN A CIT

rewriting (3) as

$$\sum (z_1, \dots, z_{N-1}) = Z_i * (N - 1), \quad (7)$$

and filling in in (6) gives:

$$Z_i * (N - 1) + z_p = 0, \quad (8)$$

$$Z_i = z_p / - (N - 1), \quad (9)$$

demonstrating the linear dependency of  $Z_i$  and  $z_p$  within one question. This linear dependency remains when the standardized responses are averaged across multiple questions.

## CHAPTER 6

### Appendix B

In this chapter, we addressed the incremental validity of the SVT above and beyond that of SCR in the detection of concealed information. The rationale here was that the SVT would yield incremental validity because it relies on a completely different cognitive mechanism than orienting, of which SCR is an index. The question whether respiration, which is also viewed as a component of the orienting reaction (Lynn, 1966), would yield incremental validity was left unanswered, however. To investigate this issue, we subjected the data of both experiment to an additional analysis. First, Respiration Line Length (RLL) was calculated as the total length of the respiration recording line in the 15 s window after stimulus onset (Timm, 1982). Next, SCR, SVT and RLL were entered as predictors in a logistic regression analysis. To identify those variables that account for an additional amount of variance, a step-by-step inclusion procedure following the Wald-statistic was used. For experiment 1, both SCR and RLL, but not SVT were included in the logistic regression model. For experiment 2, all three measures were successively included (see Tables 6.2 and 6.3). This means that in both experiments, RLL yielded incremental validity above and beyond SCR, emphasizing that the SVT should be used as an addition to, and not as a replacement of psychophysiological measures.

**Table 6.2.** Predictors of guilt in the logistic regression model of the data of two experiments reported in chapter 6.

Measure	$\beta$	S.E.	Wald	df	p
Experiment 1					
SCR	- 4.39	1.25	12.24	1	< .001
RLL	- 4.18	1.51	7.66	1	.006
(Constant)	1.44	0.51	8.13	1	.004
Experiment 2					
SVT	- 1.81	0.47	14.77	1	< .001
SCR	- 3.69	0.97	14.40	1	< .001
RLL	- 4.01	1.36	9.01	1	.003
(Constant)	4.34	0.90	19.23	1	< .001

Note. SCR = Skin Conductance Response, RLL = Respiration Line Length, SVT= Symptom Validity Test



## COMBINING SCR AND SVT IN A CIT

**Table 6.3.** Explained variance and percentage correctly classified participants for the different regression models of the data of the two experiments reported in chapter 6.

Model	Nagelkerke R <sup>2</sup>	Classification (%)
Experiment 1		
SCR	.52	77
SCR + RLL	.68	87
Experiment 2		
SVT	.58	81
SVT + SCR	.79	91
SVT + SCR + RLL	.85	94

Note. SCR = Skin Conductance Response, RLL = Respiration Line Length, SVT= Symptom Validity Test

## General discussion

This dissertation addressed the use of polygraph tests for the detection of deception. The words polygraph and lie detector are by no means synonyms. Polygraph testing includes a variety of techniques, and each technique can be applied to numerous situations. In part I, the growing use of the Control Question Test (CQT) in Europe was discussed. Chapter 1 highlighted the increased use of this technique in the legal arena. Although courts across Europe seem to be aware of the controversy and limitations that characterize the CQT, due to the 'free evidence' system that governs evidence in most countries, courts have almost never generally banned it as evidence. Chapter 2 critically examined a specific application of the CQT that is now rapidly becoming popular, namely its use in the treatment and risk assessment of convicted sex offenders. The main difference between application of the CQT in the legal arena and with sex offenders is that in the legal arena, a known incident like a murder is investigated. In sex offenders, on the other hand, it is unknown whether an incident has taken place, and questions are therefore formulated more ambiguously. It was argued that because of this difference, most CQT-research does not generalize to use with sex offenders. There is no empirical evidence for the accuracy of this particular application of the CQT, and the positive results claimed by practitioners are likely to be due to offenders disclosing behavior as a result of the intimidating effect of the polygraph rather than to its accuracy.

It is important to realize that the problem of ambiguous questions discussed in chapter 2 is not confined to sex offender testing. It plagues application of the polygraph in pre employment screening (Merckelbach & Meijer, 2008; National Research Council, 2003), supervision programs for DUI offenders (driving under the influence; Lapham, C'De Baca, Lapidus, & McMillan, 2007), but also forensic applications. Consider, for example, the following murder case:<sup>32</sup> On an early morning, a wealthy 60 year old man, is found dead in his apartment. He is found lying next to his bed with his throat cut. His wife claims that she was in bed at the time the murder took place, but did not notice anything because she was drugged by the killers. The police do not trust her story, and she agrees to take a polygraph test. During the first session, the relevant questions pertained to whether she cut her husband's throat, and test outcome was 'no deception indicated'. She was then submitted to a second session, during which the relevant

---

<sup>32</sup> This is the Belgian case of Els L. see [www.peterdevries.nl](http://www.peterdevries.nl) for details.



## GENERAL DISCUSSION

questions pertained to involvement in the crime. Now relevant questions asked were whether she knew with certainty who committed the murder of her husband, whether she helped committing this murder and whether she organized this murder. This time, test outcome was ‘deception indicated’, leading the police to believe that she did not kill her husband herself, but did order the killing. She was later sentenced to 25 years in prison. But as is true for sex offender testing, the ambiguity of the relevant questions in the second session makes comparison with the control questions cumbersome, and test outcome less accurate.

In chapter 2, it was argued that there is no solid evidence that the use of polygraph tests in sex offenders leads to decreased recidivism. A first study investigating this issue has recently been published, and deserves some attention. McGrath and colleagues (2007) compared 5-year re-offense rates of a group of 104 adult male sex offenders who received community cognitive-behavioral treatment, correctional supervision, and periodic polygraph compliance exams with a matched group of 104 sex offenders who received the same type of treatment and supervision services but no polygraph exams. In line with earlier studies, polygraph testing resulted in the disclosure of previously withheld high risk behaviors, and the vast majority of the treatment providers and supervision officers rated the tests as helpful or very helpful. Still, their data did not provide much support for an effect of polygraph tests on recidivism. Although, at fixed 5-year follow-up periods, the number of individuals in the polygraph group charged with committing a new non-sexual violent offense was significantly lower than in the no polygraph group (2.9% versus 11.5%), there were no significant between-group differences for the number of individuals charged with a new sexual (5.8% versus 6.7%), or other (non-sexual and non-violent) offense (35.6% versus 29.8%). There was also no significant difference in overall recidivism (sexual, violent, or other; 39.4% versus 34.6%). Also, the number of individuals known to have violated their supervision conditions did not differ between groups (51.9% versus 45.2%), nor did the number of individuals who returned to prison (47.1% versus 38.5%). In line with the conclusions of chapter 2, the authors argued that the polygraph’s “widespread use has far outpaced empirical examination of its effectiveness” (McGrath et al., 2007, p. 391).

In sum, part of the problems that surround the CQT can be traced back to unvalidated and improper applications like sex offender testing and personnel screening. The available evidence for the CQT’s above chance performance is limited to specific incident testing, but this application has also received much criticism. It is plagued by serious shortcomings such as a

lack of theoretical underpinning, lack of standardization and false positive rates that are beyond what any civilized criminal law system should be willing to accept.

The second part of the dissertation concerned a less widely used application of the polygraph, namely the Concealed Information Test (CIT). This test intends to determine the presence or absence of crime related information in a suspect, from which guilt can be inferred. Three different measures for the detection of concealed information were tested. Chapter 4 showed that the P300 component of the event-related potential could successfully be used for the detection of concealed face recognition. Data from a second experiment reported in this chapter further indicated that mere recognition of a face was insufficient for eliciting a P300. Chapter 5 tested the interrelationship of Skin Conductance Response (SCR) and P300 in a CIT. Results showed that SCR and P300 were uncorrelated, supporting the use of both these measures in a CIT. Finally, Chapter 6 showed that Symptom Validity Testing provides an effective means of increasing the accuracy of a SCR-based CIT procedure.

### ***Mechanisms underlying the P300 in a CIT***

The absence of a correlation between SCR and P300 in a CIT reported in chapter 5 indicates that different cognitive mechanisms underlie these measures. Further support for this notion comes from the results of the second experiment reported in chapter 4, where, in the absence of specific CIT-instructions, recognized faces did not elicit a P300. This suggests that, whereas orienting underlies autonomic responding (i.e., SCR, heart rate and respiration) in a CIT (Verschuere et al., 2004; but see Verschuere, Crombez, Koster, Van Bockstaele, & De Clercq, 2007; Gamer, Gödert, Keth, Rill, & Vossel, in press), the P300 in a CIT is due to intentional, rather than automatic processing. This contradicts the idea that the CIT is a pure passive recognition test, as was recently postulated by Meegan (2008) when he wrote that “P300 effects measure recognition rather than deception. Moreover, they can (and should) be measured without dishonest responding.” (p. 18)

That executive processes underlie deceptive responses is also a consistent finding in research using neuroimaging. In general, lying is associated with activation of executive brain regions such as prefrontal and anterior cingulate cortices (Abe, Okuda, Suzuki et al., 2008; Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003; Johnson, Barnhardt, & Zhu, 2003, 2004; Kozel, Johnson, Mu et al., 2005; Kozel, Padgett, & George, 2004; Langleben et al., 2002; Lee, Liu, Chan et al., 2005; Nunez, Casey,



## GENERAL DISCUSSION

Egner, Hare, & Hirsch, 2005; Spence, Farrow, Herford et al., 2001; Spence, Hunter, Farrow et al., 2004; Spence, Kaylor-Hughes, Brook, Lankappa, & Wilkinson, in press). These brain structures have been linked to error monitoring, showing increased activity under conditions of increased response competition and conflict (Braver, Barch, Gray, Molfese, & Snyder, 2001; Carter, Braver, Barch et al., 1998; Ruff, Woodward, Laurens, & Liddle, 2001), and response inhibition (Li, Huang, Constable, & Sinha, 2006; Li, Huang, Yan et al., 2008). This suggests that truthful responses are prepotent, and executive processes such as inhibition and conflict monitoring are essential in order to produce deceptive responses (see also Iacono, 2007; Spence et al., in press).

Indirect evidence relating these executive processes to the P300 comes from research using the stop-signal task. In this task, subjects are given a series of trials, and on each trial, a stimulus is presented that requires a response. On a minority of the trials, however, after presentation of the stimulus, some kind of “stop-signal” is presented. This stop signal instructs the subjects to inhibit their response to the primary task. ERP studies utilizing this paradigm have shown that (successful) inhibition is associated with a centrally distributed P300. (Dimoska, Johnstone, & Barry, 2006; Kok, Ramautar, De Ruiter, Band, & Ridderinkhof, 2004; Ramautar, Kok, & Ridderinkhof, 2004).<sup>33</sup> In all, it is likely that executive processes contribute to the P300 in a CIT.

The contribution of executive processes to the P300 in a CIT can also explain the finding by Rosenfeld and colleagues (2006, 2007) that autobiographical stimulus material is detected better than incidentally learned material. Of special interest here is an fMRI study by Nunez et al. (2005). These authors presented their subjects with autobiographical (e.g., do you own a laptop computer?) and nonautobiographical (e.g., is a laptop computer portable?) questions. Each question was presented twice, once with the instruction to respond truthfully, and once falsely. The results showed that deceptive responses were characterized by longer reaction times and activation of a brain circuit implicated in response conflict and

---

<sup>33</sup> This central, rather than parietal maximum can be attributed to the lack of a motor response in the stop trials. Motor responses are characterized by a centrally distributed negative waveform. In trials where a response is executed, this negative waveform dampens the P300 at Cz, leaving it maximal at Pz. On stop trials, however, no response is executed leaving the P300 maximal at Cz (Bekker, Kenemans, Hoeksma, Talsma, & Verbaten, 2005; Salisbury, Griggs, Shenton, & McCarley, 2004; Salisbury, Rutherford, Shenton, & McCarley, 2001).

## GENERAL DISCUSSION

cognitive control.<sup>34</sup> Importantly, both behavioral and neural effects were more robust when falsifying autobiographical responses relative to nonautobiographical responses, showing that the amount of conflict induced and cognitive control needed to successfully execute false responses is greater when dealing with self-relevant information. Thus, to the extent that the P300 in a CIT reflects executive processes such as inhibition, it will be increased for self-referent stimuli, which is in line with the findings by Rosenfeld and colleagues (2006, 2007).

Finally, it is worth noticing that it is unlikely that only executive processes underlie the P300 in a CIT. Two unpublished datasets from our lab show that passive recognition also plays a role. The first dataset was part of the first experiment reported in chapter 4. In the condition reported in that chapter, the participants classified all faces based on familiarity, with the instructions to deny recognition of one of the familiar faces. The results showed that the faces of which the participants denied recognition still elicited a P300. In an additional condition, not reported in chapter 4, we isolated passive recognition while ensuring that the faces were indeed processed. To achieve this, two versions of each picture were produced: one with a dot on the right cheek, and one with a dot on the left cheek (see Figure 7.1 for an example). In this dot condition, the participants were instructed to press the left or right button in correspondence to the location of the dot. P300 amplitudes for familiar and irrelevant faces for both conditions are plotted in Figure 7.2. A repeated measures ANOVA on these values with condition (face, dot) and type (probe, irrelevant) as factors revealed a significant interaction ( $F(1,23) = 13.2, p = .001$ ). Post hoc testing showed that the P300 to the probe was smaller in the condition where participants classified according to dot placement compared to classification based on familiarity ( $F(1,23) = 11.1, p = .003$ ), and no difference between the irrelevant stimuli (see Figure 7.2). Importantly, however, in the dot condition, P300 to familiar faces was still larger than to the irrelevant faces ( $F(1,23) = 35.8, p < .001$ ), showing that passive recognition was sufficient to elicit a P300.

In a similar experiment, we tested whether passive recognition of autobiographical material elicited a P300. Participants were now presented with stimuli that consisted of autobiographical information (first name, last name, father's name, mother's name, and birth date), alternated with irrelevant stimuli. Participants were instructed to respond based on the case of the stimuli. If stimuli were written in upper-case, one button was pressed,

---

<sup>34</sup> This included activation in the anterior cingulate, caudate and thalamic nuclei, and dorsolateral prefrontal cortex.

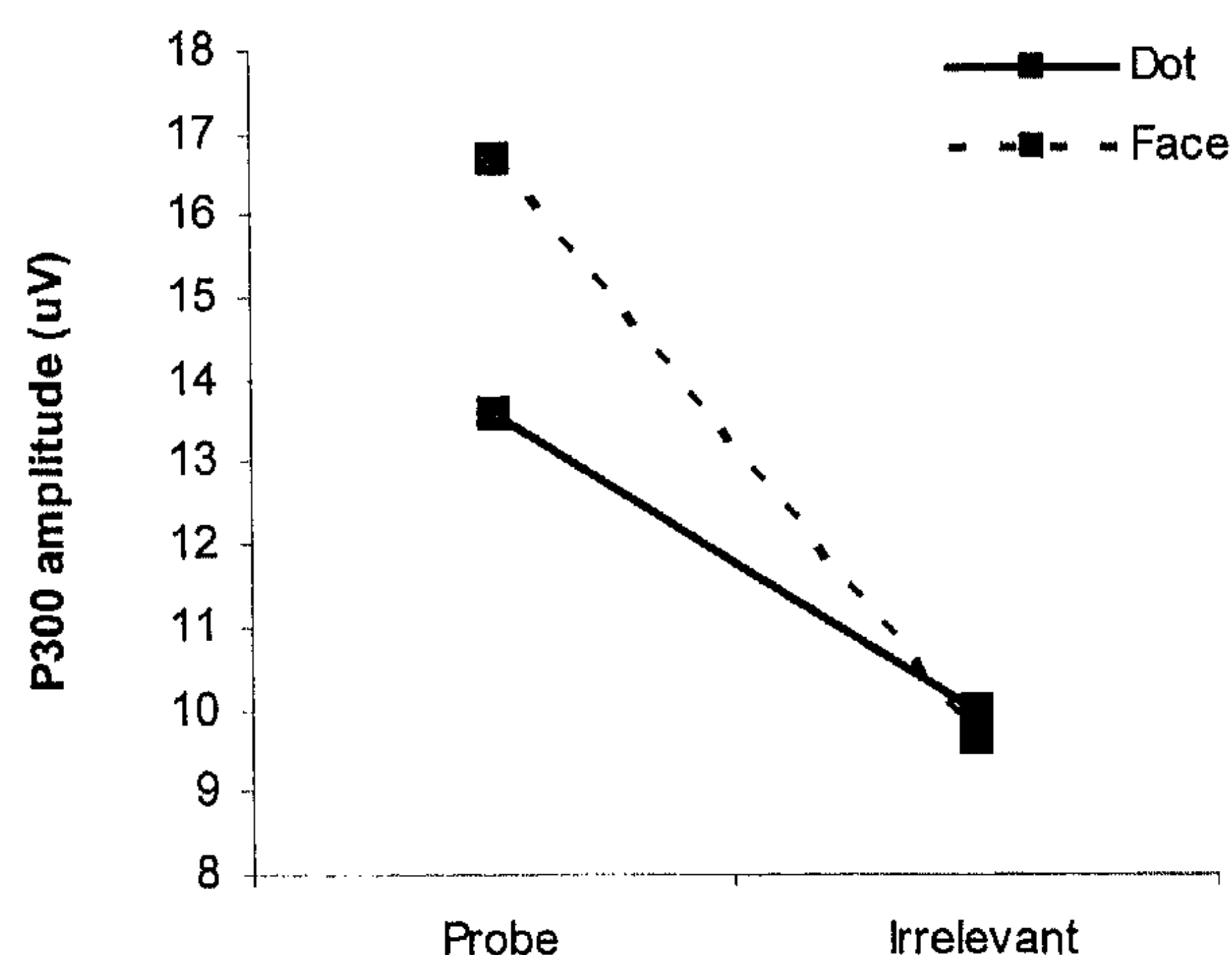


## GENERAL DISCUSSION



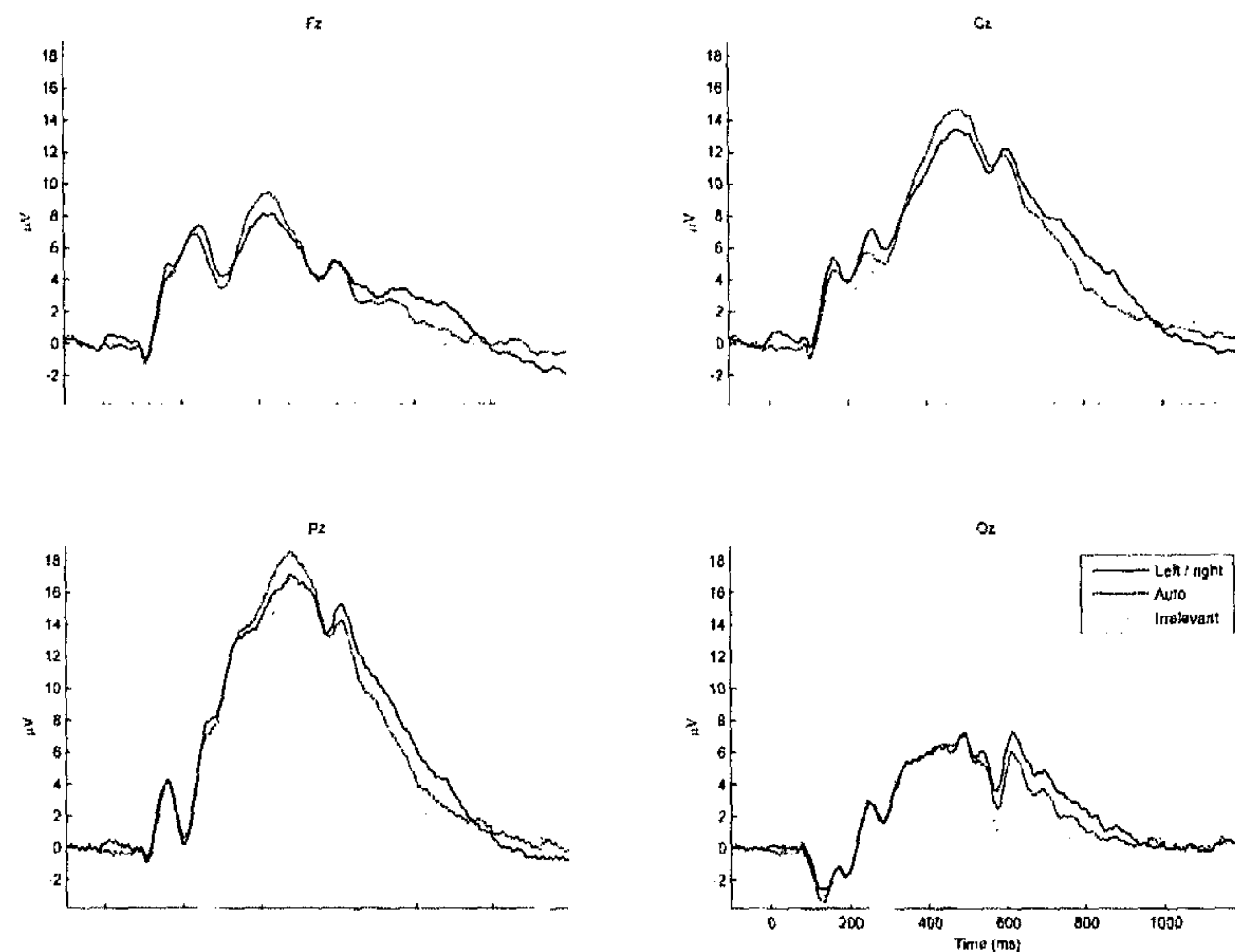
**Figure 7.1.** Example of a picture used in Experiment 1 of chapter 4 (left), with a dot on the right cheek (center), and with a dot on the left cheek (right).

whereas another was pressed for stimuli written in lower-case. To ensure processing of the stimuli, the words 'left' and 'right' were also presented, upon which the participants pressed the left and right button. Of all stimuli presented, one sixth of the stimuli were autobiographical, one sixth were the words left/right and one sixth were irrelevant stimuli. The ERP waveforms from this experiment are given in Figure 7.3. Stimuli referring to autobiographical information elicited a larger P300 than the irrelevant stimuli ( $F(1,26) = 50.9, p < .001$ ), again indicating that passive recognition elicits a P300.



**Figure 7.2.** P300 amplitude to faces classified based on familiarity (face) or dot placement (dot) for familiar faces (probes) and unfamiliar faces (irrelevants).

## GENERAL DISCUSSION



**Figure 7.3.** Grand average ERPs at Fz, Cz, Pz, and Oz as a function of stimulus type (left / right: word left or right; Auto: autobiographical stimuli; irrelevant: irrelevant information).

In sum, both recognition and executive processes are likely to contribute to the P300 in a CIT. Still, in both datasets reported above, passive recognition was sufficient to elicit a P300, whereas in experiment 2 reported in chapter 4 it was not. It is important to note that the two datasets reported above used highly significant stimulus material. Experiment 2 of chapter 4, on the other hand, combined the use of less significant stimuli (faces of university professors) with the absence of specific task instructions to deny recognition. It may be that it was this combination was responsible for the lack of a P300.

For practical purposes, the P300 seems to possess only limited additional value over and beyond measures of the autonomic nervous system like SCR. Measuring P300 in a CIT requires extensive preparation and a technical data analysis. Therefore, serial testing is recommended, with the SCR based CIT administered first. Under special circumstances, for example in non-responding suspects, or under the suspicion of a false negative outcome, this can be followed by a P300 based CIT. Given the positive relationship between psychopathic traits and the detectability with a



## GENERAL DISCUSSION

prevalence of such claims are given by Cima and colleagues (Cima, Nijman, Merckelbach, Kremer, & Hollnack, 2004). These researchers analyzed the file records of 308 German and Dutch forensic patients, and showed that 23% reported partial or complete amnesia for their crimes. Organic factors such as intoxication or brain damage may account for severe memory loss, but given the strategic advantages that claims of memory loss can have, simulation should also be considered. Within the legal community, there is a consensus that amnesia is easily faked, often occurs, and is hard to detect (Christianson, Merckelbach, & Kopelman, 2006; Tysse, 2005; Wiggins & Brandt, 1988). Therefore, in such cases, the CIT may prove a helpful assistant in unmasking malingered memory loss.

### *fMRI and the detection of deception*

One of the more recent techniques for the detection of deception that has received much attention is the measurement of brain activation with fMRI. This method is currently commercialized by a company with the name 'No Lie MRI',<sup>35</sup> and has received much attention, not the least from the (popular scientific) press. Whether fMRI based lie detection will ever become applicable in the field remains doubtful, however, for a number of reasons.

First, there is the issue of reverse inference (Poldrack, 2006). We saw earlier that lying is associated with more activation of executive brain regions such as prefrontal and anterior cingulate cortices than truth telling. These executive processes are, however, not uniquely related to deception. Moreover, different types of lies are associated with different processes and hence activation in different areas (Ganis et al., 2003). It is highly unlikely that deception relies on a single cognitive process or function, which also explains the differences between studies (Spence, 2008). This means that although deception is associated with some aspects of executive processes, finding correlates of these executive processes on an fMRI scan by no means ensures that the individual is lying. This problem is very similar to the problem that plagues the CQT; lying is associated with stress, but this does not mean that recording stress in an individual means he or she is lying. This means that, as with SCR and P300, the practical merit of fMRI may be limited to the detection of concealed information, rather than the detection of lies.

---

<sup>35</sup> On their website ([www.noliemri.com](http://www.noliemri.com)), it is claimed that "No Lie MRI™ is a proprietary product that objectively measures intent, prior knowledge, and deception using algorithms to automatically analyze functional Magnetic Resonance Imaging (fMRI)." And: "The approach used by No Lie MRI™ has a verified accuracy that greatly surpasses all other truth verification / lie detection methods."

## GENERAL DISCUSSION

Another issue that currently hinders practical application of fMRI based lie detection is a lack of validated scoring algorithms yielding decisions at the individual level. The patterns of activation observed during deceit emerge when averaging over all participants. Using fMRI for lie detection, however, requires a decision at the individual level. This can be problematic, because the variability between individuals is large (Miller, Van Horn, Wolford et al., 2002), and as a consequence, classification rates will be far from perfect. So far, only three studies have tried individual classification (Davatzikos, Ruparel, Fan et al., 2005; Kozel et al., 2005; Langleben, Loughead, Bilker et al., 2005), finding correct classification rates between 78% and 90%. How well these rates generalize to other paradigms or datasets, however, remains to be seen.

The current evidence on fMRI based lie detection does not justify practical application (see also Sip, Roepstorff, McGregor, & Frith, 2008; Spence, 2008). The high expectations ascribed to fMRI based lie detection by lay people can probably be attributed to factors other than scientific evidence. Germane to this issue is a study by Weisberg and co-workers (2008), that showed how easily lay people are satisfied by scientific-sounding, but in fact meaningless neuroscientific evidence. These authors presented their participants with psychological phenomena like “the curse of knowledge”.<sup>36</sup> Next, participants were given either a good or a bad explanation for each phenomenon. The good explanations were the genuine explanations that the researchers gave (e.g., researchers claim this “curse” happens because subjects have trouble switching their point of view to consider what someone else might know, mistakenly projecting their own knowledge onto others.) The bad explanations were circular restatements of the phenomenon, hence, not explanatory (e.g., the researchers claim that this “curse” happens because subjects make more mistakes when they have to judge the knowledge of others. People are much better at judging what they themselves know.) Three groups of participants - neuroscience experts, novices, and students - were asked to rate how satisfying they found good

---

<sup>36</sup> This curse of knowledge was explained to the participants as follows: Researchers created a list of facts that about 50% of people knew. Subjects in this experiment read the list of facts and had to say which ones they knew. They then had to judge what percentage of other people would know those facts. Researchers found that the subjects responded differently about other people’s knowledge of a fact when the subjects themselves knew that fact. If the subjects did know a fact, they said that an inaccurately large percentage of others would know it, too. For example, if a subject already knew that Hartford was the capital of Connecticut, that subject might say that 80% of people would know this, even though the correct answer is 50%. The researchers call this finding “the curse of knowledge.” (Weisberg et al., 2008, p. 471)



## GENERAL DISCUSSION

and bad explanations. They all recognized the difference between the good and the bad explanations, and rated the good explanations with a positive score (satisfying) and the bad explanation with a negative score (unsatisfying). The researchers then repeated the whole procedure with new participants, but this time, irrelevant neuroscience information was added to the beginning of both explanations such as: “Brain scans indicate that this “curse” happens because of the frontal lobe brain circuitry known to be involved in self-knowledge.” The experts recognized the insufficiency of the neuroscience information, and now rated both the good and the bad explanations as unsatisfying. Striking, however, was the rating of the two groups with novices and students. They rated the bad explanations as much more satisfying now that they were preceded by trivial neuroscience, showing the impact of scientific-sounding but empirically and conceptually uninformative neuroscientific information on lay people.

In sum, at a theoretical level, fMRI research has resulted in an important body of knowledge about psychological processes underling deception. With this, it has brought about important progress in the field addressing the criticism echoed by the National Research Council that polygraph research has failed to build and refine its theoretical base. The field is, however, also characterized by serious interpretational problems, of which some are of a logical nature. The promise that fMRI based lie detection is ascribed by lay people, can probably be better explained by the neuroscientific connotations, than by neuroscientific evidence.

### *Voice analysis and SCAN*

The great need for accurate lie detection methods has also attracted the attention and involvement of nonacademic parties. These parties market lie detection techniques which they claim are highly accurate and viable alternatives to the polygraph. These methods are widely used, regardless of the absence of scientific evidence for the diagnostic accuracy of any of these methods, as we will argue below.

Attempts to infer deception from the voice have been around since the sixties (Lykken, 1998). The devices that are nowadays marketed have names like *Computer Voice Stress Analyzer*, *Layered Voice Analysis*, *Security Call Analyzer* and *Gate Keeper*. These devices all rely on the assumption that lying is reliably associated with changes in the voice, and that these changes can

## GENERAL DISCUSSION

be picked up through voice analysis (see also de Paulo et al., 2003).<sup>37</sup> If these devices would indeed be capable of detecting deceit, their advantages are obvious. They would be useful at airports, for job interviews, and could be widely used by insurance companies and law enforcement organizations. Unfortunately, however, empirical support for the effectiveness of these devices is lacking (Gamer, Rill, Vossel, & Godert, 2006; Krapohl, Ryan, & Shull, 2002; Lykken, 1998; National Research Council, 2003). Still, the manufacturers seem to have great confidence in their products. For example, recently, a sales representative contacted our research group, asking whether we were interested in conducting scientific research with their product. After all, so he claimed, their device could achieve an accuracy of 90%.

We tested their \$10,000 voice analysis software, with disappointing results. For a field test, we used a video taped interview by Dutch crime reporter Peter R. de Vries. The interviewee had moved into a house belonging to a couple that had been missing for several years. According to him the couple had moved to Belgium. They occasionally dropped by, and he was taking care of the house in the meantime. He was later convicted for murdering the couple, and their bodies were found at a children's farm where he worked on a voluntary basis. During the interview several statements were made, of which it could be determined that they were outright lies. For example, the interviewee said he had no access to the bank accounts of the couple, whereas bank files showed otherwise. As we fed the tape to the voice analysis software, its diagnostic classifications ranged from ambiguous terms such as 'exited' to 'subject not sure'. But nowhere during the interview did it pick up on any of the outright lies by classifying them as a 'probable lie' or 'deception'.

For a test in a more controlled environment, we moved to the laboratory. Twenty-four volunteers were asked to write down an emotional event they had experienced personally in approximately 400 words. Next, each volunteer was given one of the other volunteers' written statements, with the instruction to read it carefully so they could later reproduce it in front of the camera. Then, each volunteer was videotaped telling their own authentic story, as well as a story they themselves had not experienced. This way, 24 truthful and 24 deceptive video statements were created, all of which were fed to the voice analysis software. The results were disappointing again: the average 'global honesty rating' the software came

---

37 The manufacturer of the software we tested explained the rationale as follows: "the software uses a patented and unique technology to detect "traces" of brain activity using the voice as a medium. The technology is based on the idea that changes in cortical perception and interpretation of events manifest themselves in the vocal waveform during speech."



## GENERAL DISCUSSION

up with did not differ between the deceptive and the truthful stories (Merckelbach & Meijer, 2007). In sum, we found there to be little truth to the representative's 90% accuracy estimate.

Another method that is marketed as extremely effective in the detection of deception is Scientific Content Analysis (SCAN), developed by the Israeli criminologist, former police lieutenant, and polygrapher Avinoman Sapir. It is currently employed by police and military in countries like the United States, Canada, Belgium, Israel, Mexico, Singapore, South Africa, the Netherlands, and the United Kingdom (Vrij, 2008). The idea behind this method is that a statement derived from memory of an actual experience differs in content and quality from a statement based on invention of fantasy. Consequently, truthful statements are thought to have different linguistic and structural features than deceptive statements. SCAN requires a written statement, which can be from a defendant, witness or alleged victim. The presence of certain elements in this statement is seen as a sign of deception, and can guide the subsequent interrogation.

Large scale systematic research about SCAN is absent, and only three studies have been published. One of these three concerned a mock crime study (Porter & Yuille, 1995). The authors of that study found that none of the three SCAN criteria they tested discriminated between truthful and fabricated statements. Diagnostic accuracy estimates from the two available field studies are more positive, but both lack independent ground truth. Therefore, they should be interpreted with great caution. Driscoll (1994) analyzed 30 written statements, and SCAN correctly classified 73% of the truthful and 95% of the deceptive statements. Smith (2001), finally, had five teams of detectives judge 27 statements. Three of these teams were SCAN trained detectives with varying degrees of experience. The other two groups consisted of detectives without SCAN training. One group consisted of experienced and the other of inexperienced detectives. All five groups performed above chance level. The three SCAN groups did not differ from each other and classified at least 80% of the true statements and 75% of the false statements correctly. Similar classification rates were achieved by the experienced detectives without SCAN, however, making SCAN no better than experienced detectives.

To conclude, the empirical and theoretical basis for SCAN is weak, at best. Many of the elements supposedly indicating deception lack empirical support, and it remains unclear how they relate to the theoretical framework. Even though research offers some support for differences in linguistic style between truthful and deceptive statements included in SCAN (e.g., deceptive statements contain fewer self references; Newman,

## GENERAL DISCUSSION

Pennebaker, Berry, & Richards, 2003; but see Bond & Lee, 2005), its widespread use has also far outpaced empirical examination of its effectiveness.

### *Future Research*

In chapter 2, we argued that the admissions made by sex offenders were a result of the polygraph's intimidating effect rather than its accuracy (bogus pipeline; see also Gannon et al., 2007; Lykken, 1998; Roese & Jamieson, 1993). The exact reason as to why the polygraph is perceived as intimidating has, however, received little attention. A concept that might be of interest to this issue is that of interoceptive awareness. This term refers to an individuals' ability to perceive bodily signals and their changes, and has played an important role in many theories of emotion (Damasio, 1999; James, 1884). These theories state that the perception of bodily states is a crucial determinant for the processing and the subjective experience of emotions and feelings (Pollatos, Gramann, & Schandry, 2007). Importantly, there are differences in the accuracy by which individuals can perceive their own bodily signals. Moreover, individuals with more accurate perception of their bodily signals show larger physiological responses to emotional stimuli (Pollatos, Herbert, Matthias, & Schandry, 2007; Suchday & Larkin, 2004; Wiens, Mezzacappa, & Katkin, 2000), and experience more intense emotions (Craig, 2004; Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004; Pollatos et al., 2007). One avenue for future research could entail looking at whether this interoceptive awareness is related to polygraph induced confessions. After all, the crucial questions in a CQT test are easily recognizable, and it would be interesting to see to what extent perception of the bodily signals and the concomitant intense emotions to these questions influence a suspect decision to confess.

With regard to the CIT, there are a number of questions that future research should focus on. Firstly, research should focus on what constitutes good CIT material. Unanswered questions that currently hinder application of the CIT, for example, pertain to what information is encoded under high stress situations (see e.g., Morgan III, Hazlett, Doran et al., 2004) and how long this information is retained in memory. This type of research could be conducted either using ecologically valid manipulations, or by interviewing perpetrators about the crimes they have committed.

Secondly, future research should aim at further advancing our insights into the psychological mechanisms underlying the CIT. This can be done by employing a variety of measures, including, but not limited to brain imaging. Expanding our knowledge about underlying mechanisms will have



## GENERAL DISCUSSION

important practical implications. To the extent that, for example, response conflict underlies the P300 in a CIT, the use of highly similar stimuli will maximize detection efficiency, because these stimuli induce a response conflict. This, however, would be at odds with recommendations that stimuli should be used that are easily perceived as different, so as to minimize response generalization across similar stimuli (Iacono, 2007).

Furthermore, efforts should be made to apply the CIT in the field. Almost 50 years of research has resulted in a test with scientific underpinnings that are strong enough for passing the Daubert standard (Ben Shakhar et al., 2002; Iacono, 2008).<sup>38</sup> Nevertheless, no western country has been able to implement the test, not even on a small scale. It is unlikely that investigative authorities will successfully implement the method on their own. Therefore, researchers should be willing to invest in assisting the investigative authorities with setting up an infrastructure that allows for conducting CITs. Because of the low probability of false positive errors, the CIT can be tested under field situations relatively safely. After all, the risk of incriminating an innocent suspect is small. This type of research will yield important field data that is currently largely absent.

Finally, when considering field application of the CIT, the use of countermeasures is of importance. Countermeasures refer to anything the examinee may do to defeat or distort a polygraph examination (Honts & Amato, 2002). Having the participants perform specific actions upon presentation of the different answer alternatives, for example, has been shown to reduce the sensitivity of CITs based on skin conductance (Ben-Shakhar & Dolev, 1996; Honts et al., 1996) and P300 (Mertens & Allen, 2008; Rosenfeld et al., 2004). Also, supplying participants with the rationale of the SVT has been shown to reduce its effectiveness (Verschuere, Meijer, & Crombez, in press). Future research could look at whether a combination of SCR, P300 and SVT is equally vulnerable to countermeasures. After all, the data presented in this dissertation suggest that they rely on different psychological mechanisms, and they may therefore not be susceptible to the same countermeasures.

---

<sup>38</sup> This standard is provided by the United States Supreme Court for admitting scientific evidence in federal courts. It constitutes of the following criteria (see e.g., Ben Shakhar et al., 2002): 1) is the evidence based on a testable theory or technique; 2) has the theory or technique been peer reviewed; 3) does it have a known error rate; and 4) is the underlying science generally accepted within the relevant scientific community?

## GENERAL DISCUSSION

### *Concluding remarks*

In this dissertation, a variety of techniques and dependent measures for the detection of deception have been critically reviewed. Current developments in Europe are worrying, as techniques without scientific standing like SCAN and sex offender polygraph testing are increasingly being used. Testing for concealed information, on the other hand, is substantiated by enough empirical data, and offers sufficient protection of innocent examinees to warrant field application. It can be done with a range of dependent measures such as cost effective and easily implemented forced choice procedures, autonomous measures like skin conductance and respiration, but also with measures of brain activity. If several practical hurdles are overcome, the CIT will be a valuable addition to the investigative authorities' toolkit.



## References

- Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matsuda, T., Mori, E., et al. (2008). Neural Correlates of True Memory, False Memory, and Deception. *Cerebral Cortex*.
- Abrams, S., & Abrams, J. B. (1993). *Polygraph testing of the pedophile*. Portland: Ryan Gwinner Press.
- Ahlmeier, S., Heil, P., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 12, 123-138.
- Allen, J. J., & Iacono, W. G. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology*, 34, 234-240.
- Allen, J. J., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology*, 29, 504-522.
- Andreassi, J. L. (2000). *Psychophysiology: Human behavior and physiological response* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Arnett, P. A. (1997). Autonomic responsivity in psychopaths: A critical review and theoretical proposal. *Clinical Psychology Review*, 17, 903-936.
- Babiak, P., & Hare, R. D. (2006). *Snakes in suits: When psychopaths go to work*. New York: Regan Books / Harper Collins Publishers.
- Bahramali, H., Gordon, E., Lim, C. L., Li, W., Lagopoulos, J., Leslie, J., et al. (1997). Evoked related potentials associated with and without an orienting reflex. *Neuroreport*, 8, 2665-2669.
- Balloun, K. D., & Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology*, 64, 316-322.
- Barry, R. J., Feldmann, S., Gordon, E., Cocker, K. I., & Rennie, C. (1993). Elicitation and habituation of the electrodermal orienting response in a short interstimulus interval paradigm. *International Journal of Psychophysiology*, 15, 247-253.
- Begleiter, H., Porjesz, B., & Wang, W. (1995). Event-related brain potentials differentiate priming and recognition to familiar and unfamiliar faces. *Electroencephalography and Clinical Neurophysiology*, 94, 41-49.
- Bekker, E. M., Kenemans, J. L., Hoeksma, M. R., Talsma, D., & Verbaten, M. N. (2005). The pure electrophysiology of stopping. *International Journal of Psychophysiology*, 55, 191-198.
- Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, 22, 292-299.
- Ben-Shakhar, G. (1991). Future prospects of psychophysiological detection: Replacing the CQT by the GKT. In J. R. Jennings, P. K. Ackles & M. G. H. Coles (Eds.), *Advances in Psychophysiology* (Vol. 4, pp. 193-199). London: Jessica Kingsley.
- Ben-Shakhar, G. (in press). The case against the use of polygraph examinations to monitor post-conviction sex offenders. *Legal and Criminological Psychology*.
- Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology*, 81, 273-281.

## REFERENCES

- Ben-Shakhar, G., & Elaad, E. (2002). Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: A reexamination. *Journal of Applied Psychology, 87*, 972-977.
- Ben Shakhar, G. (2002). A critical review of the Control Questions Test (CQT). In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 103-126). San Diego: Academic Press.
- Ben Shakhar, G., Bar Hillel, M., & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the use of the guilty knowledge technique in court. *Law and Human Behavior, 26*, 527-541.
- Ben Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology, 88*, 131-151.
- Ben Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective*. New York: Springer-Verlag Publishing.
- Ben Shakhar, G., Gronau, N., & Elaad, E. (1999). Leakage of relevant information to innocent examinees in the GKT: An attempt to reduce false-positive outcomes by introducing target stimuli. *Journal of Applied Psychology, 84*, 651-660.
- Benning, S. D., Patrick, C. J., Hicks, B. M., Blonigen, D. M., & Krueger, R. F. (2003). Factor structure of the psychopathic personality inventory: Validity and implications for clinical assessment. *Psychological Assessment, 15*, 340-350.
- Bennington, J. Y., & Polich, J. (1999). Comparison of P300 from passive and active tasks for auditory and visual stimuli. *International Journal of Psychophysiology, 34*, 171-177.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*, 551-565.
- Berlad, I., & Pratt, H. (1995). P300 in response to the subject's own name. *Electroencephalography and Clinical Neurophysiology, 96*, 472-474.
- Blackstone, W. (1882). *Commentaries on the laws of England* (3rd ed.). London: Murray.
- Blair, J., Mitchel, D., & Blair, K. (2005). *The Psychopath: Emotions and the brain*. Oxford: Blackwell Publishing.
- Bobes, M. A., Lopera, F., Comas, L. D., Galan, L., Carbonell, F., Bringas, M. L., et al. (2004). Brain potentials reflect residual face processing in a case of prosopagnosia. *Cognitive Neuropsychology, 21*, 691-718.
- Bockstaele, M. (2000). *De polygraaf*. Brussel: Politeia.
- Boelhouwer, A. J. W., Merckelbach, H., van Koppen, P. J., & Verbaten, M. N. (1996). *Leugendetecie in Nederland* (No. Rapport aangeboden aan de minister van Justitie).
- Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19*, 313-329.
- BPS Working Party. (2004). *A review of the current scientific status and fields of application of polygraph deception detection*: British Psychological Society.
- Brady, J. P., & Lind, D. L. (1961). Experimental analysis of hysterical blindness. *Archives of General Psychiatry, 4*, 331-339.
- Branaman, T. F., & Gallagher, S. N. (2005). Polygraph testing in sex offender treatment: A review of limitations. *American Journal of Forensic Psychology, 23*, 45-64.
- Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex, 11*, 825-836.
- Breen, N., Caine, D., & Coltheart, M. (2000). Models of face recognition and delusional misidentification: A critical review. *Cognitive Neuropsychology, 17*, 55-71.



## REFERENCES

- British Psychological Society. (1996). Report of the Working Group on the use of the polygraph in criminal investigation and personnel screening. *Bulletin of the British Psychological Society*, 39, 81-94.
- Bumby, K. M. (1996). Assessing the cognitive distortions in of child molesters and rapists: Developments and validation of the MOLEST and RAPE scales. *Sexual Abuse: A Journal of Research and Treatment*, 8, 37-54.
- Buracas, G. T., & Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage*, 16, 801-813.
- Bureau of Justice Statistics. (2003). *Recidivism of sex offenders released from prison in 1994*. Washington, DC: U.S. Department of Justice.
- Burgsmüller, C. (2000). Das BGH-Urteil zu den Glaubhaftigkeitsgutachten: Eine späte Folge der sog. Wormser Strafverfahren vor dem Landgericht Mainz? *Praxis der Rechtspsychologie*, 10, 48-59.
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: The external validity of mock crime studies. *Journal of Experimental Psychology: Applied*, 9, 261-269.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280, 747-749.
- Castro, A., & Diaz, F. (2001). Effect of the relevance and position of the target stimuli on P300 and reaction time. *International Journal of Psychophysiology*, 41, 43-52.
- Christianson, S., & Merckelbach, H. (2004). Crime-related amnesia as a form of deception. In P. A. Granhag & L. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 195-228). Cambridge: Cambridge University Press.
- Christianson, S., Merckelbach, H., & Kopelman, M. (2006). Crime-related amnesia. In A. Heaton-Armstrong, E. Shepherd, G. H. Gudjonsson & D. Wolchover (Eds.), *Witness Testimony. Psychological, Investigative and Evidential Perspectives*. Oxford: Oxford University Press.
- Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R., et al. (2003). The other side of malingering: Supernormality. *Clinical Neuropsychology*, 17, 235-243.
- Cima, M., Nijman, H., Merckelbach, H., Kremer, K., & Hollnack, S. (2004). Claims of crime-related amnesia in forensic patients. *International Journal of Law and Psychiatry*, 27, 215-221.
- Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Consigli, C. J. E. (2002). Post-conviction sex offender testing and the American Polygraph Association. In M. Kleiner (Ed.), *Handbook of polygraph testing*. (pp. 237-250). San Diego, CA: Academic Press.
- Craig, A. D. (2004). Human feelings: why are some more aware than others? *Trends in Cognitive Sciences*, 8, 239-241.
- Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7, 189-195.
- Cross, T. P., & Saxe, L. (1992). A critique of the validity of polygraph testing in child abuse cases. *Journal of Child Sexual Abuse*, 1, 19-33.
- Cross, T. P., & Saxe, L. (2001). Polygraph testing and sexual abuse: The lure of the magic lasso. *Child Maltreatment*, 6, 195-206.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: John Wiley.
- Damasio, A. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Heinemann.

## REFERENCES

- Damaška, M. R. (1986). *The faces of justice and state authority: A comparative approach to the legal process*. New Haven, CO: Yale University Press.
- Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., et al. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, 28, 663-668.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.
- Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 200-223). Cambridge: University Press.
- Denney, R. L. (1996). Symptom Validity Testing of remote memory in a criminal forensic setting. *Archives of Clinical Neuropsychology*, 11, 589-603.
- Dimoska, A., Johnstone, S. J., & Barry, R. J. (2006). The auditory-evoked N2 and P3 components in the stop-signal task: indices of inhibition, response-conflict or error-detection? *Brain and Cognition*, 62, 98-112.
- Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral Brain Sciences*, 11, 357-374.
- Doob, A. N., & Kirschenbaum, H. M. (1973). Bias in police lineups-partial remembering. *Journal of Police Science and Administration*, 1, 187-293.
- Driscoll, L. N. (1994). A validity assessment of written statements from suspects in criminal investigations using the SCAN technique. *Police Studies*, 17, 77-88.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, 75, 521-529.
- Elaad, E. (1998). The challenge of the Concealed Knowledge Polygraph Test. *Expert Evidence*, 6, 161-187.
- Elaad, E. (2003). Is the inference rule of the "control question polygraph technique" plausible? *Psychology, Crime and Law*, 9, 37-47.
- Elaad, E., & Ben-Shakhar, G. (2006). Finger pulse waveform length in the detection of concealed information. *International Journal of Psychophysiology*, 61, 226-234.
- Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, 77, 757-767.
- Ellis, H. D., & Lewis, M. B. (2001). Capgras delusion: A window on face recognition. *Trends in Cognitive Sciences*, 5, 149-156.
- Ellwanger, J., Rosenfeld, J. P., Hankin, B. L., & Sweet, J. J. (1999). P300 as an index of recognition in a standard and difficult match-to-sample test: A model of amnesia in normal adults. *Clinical Neuropsychology*, 13, 100-108.
- Ellwanger, J., Rosenfeld, J. P., Sweet, J. J., & Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *International Journal of Psychophysiology*, 23, 9-23.
- Emerick, R. L., & Dutton, W. A. (1993). The effect of polygraphy on the self report of adolescent sex offender: Implications for risk assesment. *Annals of Sex Research*, 6, 83-103.
- English, K., Jones, L., Patrick, D., & Pasini-Hill, D. (2003). Sexual offender containment: Use of the postconviction polygraph. *Annals New York Academy of Sciences*, 989, 411-427.
- Faller, K. C. (1997). The polygraph, its use in cases of alleged sexual abuse: An exploratory study. *Child Abuse and Neglect*, 21, 993-1008.
- Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, 28, 531-547.



## REFERENCES

- Fiedler, K. (1999). Gutachterliche Stellungnahme zur wissenschaftlichen Grundlage der Lügen-detektion mit Hilfe sogenannter Polygraphentests. *Praxis der Rechtspsychologie*, 9, 5-44.
- Fiedler, K., Schmid, J., & Stahl, T. (2002). What is the current truth about polygraph lie detection. *Basic and Applied Social Psychology*, 24, 313-324.
- Frederick, R. I., Carter, M., & Powel, J. (1995). Adapting symptom validity testing to evaluate suspicious complaints of amnesia in medicolegal evaluations. *Bulletin of the American Academy of Psychiatry and the Law*, 23, 227-233.
- Friedman, D., & Johnson, R., Jr. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research Technique*, 51, 6-28.
- Fukuda, K. (2001). Eye blinks: New indices for the detection of deception. *International Journal of Psychophysiology*, 40, 239-245.
- Furedy, J. J. (1996a). The North American polygraph and psychophysiology: Disinterested, uninterested, and interested perspectives. *International Journal of Psychophysiology*, 21, 97-105.
- Furedy, J. J. (1996b). Some elementary distinctions among, and comments concerning, the 'control' question 'test' (CQT) polygrapher's many problems: A reply to Honts, Kircher and Raskin. *International Journal of Psychophysiology*, 22, 53-59.
- Gamer, M., Gödert, H. W., Keth, A., Rill, H. G., & Vossel, G. (in press). Electrodermal and phasic heart rate responses in the Guilty Actions Test: Comparing guilty examinees to informed and uninformed innocents. *International Journal of Psychophysiology*.
- Gamer, M., Rill, H. G., Vossel, G., & Godert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *International Journal of Psychophysiology*, 60, 76-87.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cerebral Cortex*, 13, 830-836.
- Gannon, T. A., Keown, K., & Polaschek, D. L. (2007). Increasing honest responding on cognitive distortions in child molesters: The bogus pipeline revisited. *Sex Abuse*, 19, 5-22.
- Gannon, T. A., & Polaschek, D. L. L. (2006). Cognitive distortions in child molesters: A re-examination of key theories and research. *Clinical Psychology Review*, 26, 1000-1019.
- Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology*, 40, 216-224.
- Gronau, N., Ben-Shakhar, G., & Cohen, A. (2005). Behavioral and physiological measures in the detection of concealed information. *Journal of Applied Psychology*, 90, 147-158.
- Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19-30.
- Grubin, D., & Madsen, L. (2006). Accuracy and utility of post-conviction polygraph testing of sex offenders. *British Journal of Psychiatry*, 188, 479-483.
- Grubin, D., Madsen, L., Parsons, S., Sosnowski, D., & Warberg, B. (2004). A prospective study of the impact of polygraphy on high-risk behaviors in adult sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 16, 209-222.

## REFERENCES

- Gudjonsson, G. H. (1982). Some psychological determinants of electrodermal responses to deception. *Personality and Individual Differences*, 3, 381-391.
- Hagler, H. L. (1995). Polygraph as a measure of progress in the assessment, treatment, and surveillance of sex offenders. *Sexual Addiction and Compulsivity*, 2, 98-111.
- Hall, G. C. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology*, 63, 802-809.
- Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148, 839-843.
- Hanson, R. K. (2003). Who is dangerous and when are they safe? Risk assessment with sexual offenders. In B. J. Winick & J. Q. La Fond (Eds.), *Protecting society from sexually dangerous offenders: Law, justice, and therapy* (pp. 63-74). Washington, DC: American Psychological Association.
- Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348-362.
- Hanson, R. K., Gordon, A., Harris, A. J., Marques, J. K., Murphy, W., Quinsey, V. L., et al. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sex Abuse*, 14, 169-194; discussion 195-167.
- Hare, R. (1999). *Without conscience: The disturbing world of the psychopaths among us*. New York: The Guilford Press.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15, 413-425.
- Haughton, P. M., Lewsley, A., Wilson, M., & Williams, R. G. (1979). A forced-choice procedure to detect feigned or exaggerated hearing loss. *British Journal of Audiology*, 13, 135-138.
- Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., & Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cerebral Cortex*, 13, 793-805.
- Heydon, J. D. (2004). *Cross on evidence* (8th ed.). London: Butterworths.
- Hines, T. (2003). *Pseudoscience and the paranormal*. Amherst, NY: Prometheus books.
- Hira, S., & Furumitsu, I. (2002). Polygraphic examinations in Japan: Application of the guilty knowledge test in forensic investigations. *International Journal of Police Science & Management*, 4, 16-27.
- Hirota, A., Sawada, Y., Tanaka, G., Nagano, Y., Matsuda, I., & Takasawa, N. (2003). A new index for psychophysiological detection of deception: Applicability of normalized pulse volume. *Japanese Journal of Physiological Psychology and Psychophysiology*, 21, 217-230.
- Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, 11, 967-974.
- Honts, C. R. (2004). The psychophysiological detection of deception. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 103-126). Cambridge: Cambridge University Press.
- Honts, C. R., & Amato, S. L. (2002). Countermeasures. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 251-264). San Diego, CA: Academic Press.
- Honts, C. R., Devitt, M. K., Winbush, M., & Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, 33, 84-92.



## REFERENCES

- Honts, C. R., Kircher, J. C., & Raskin, D. C. (1995). Polygrapher's dilemma or psychologist's chimaera: A reply to Furedy's logico-ethical considerations for psychophysiological practitioners and researchers. *International Journal of Psychophysiology*, 20, 199-207.
- Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests. In J. R. Jennings, P. K. Ackles & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 4, pp. 201-207). London: Jessica Kingsley Publishers.
- Iacono, W. G. (2007). Detection of deception. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 688-703). New York: Cambridge University Press.
- Iacono, W. G. (2008). The forensic application of "brain fingerprinting:" why scientists should encourage the use of P300 memory detection methods. *American Journal of Bioethics*, 8, 30-32.
- Ito, T., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, 36, 660-676.
- James, W. (1884). What is emotion? *Mind*, 9, 188-205.
- Jelicic, M., Merckelbach, H., Timmermans, M., & Candel, I. (2004). De Nederlandstalige versie van de Psychopathic Personality Inventory Psychodiagnostisch gereedschap [The Dutch version of the Psychopathic Personality Inventory: Some psychometric results]. *De Psycholoog*, 39, 604-608.
- Jelicic, M., Merckelbach, H., & van Bergen, S. (2004a). Symptom validity testing of feigned crime-related amnesia: A simulation study. *The Journal of Credibility Assessment and Witness Psychology*, 5, 1-8.
- Jelicic, M., Merckelbach, H., & Van Bergen, S. (2004b). Symptom validity testing of feigned amnesia for a mock crime. *Archives of Clinical Neuropsychology*, 19, 525-531.
- Johnson, M. M., & Rosenfeld, J. P. (1992). Oddball-evoked P300-based method of deception detection in the laboratory. II: Utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology*, 12, 289-306.
- Johnson, R., Barnhardt, J., & Zhu, J. (2003). The deceptive response: effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biological Psychology*, 64, 217-253.
- Johnson, R., Jr. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, 30, 90-97.
- Johnson, R., Jr., Barnhardt, J., & Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia*, 42, 878-901.
- Jokinen, A., Santtila, P., Ravaja, N., & Puttonen, S. (2006). Salience of guilty knowledge test items affects accuracy in realistic mock crimes. *International Journal of Psychophysiology*.
- Joyce, C. A., & Kutas, M. (2005). Event-related potential correlates of long-term memory for briefly presented faces. *Journal of Cognitive Neuroscience*, 17, 757-767.
- Jutai, J. W., Hare, R. D., & Connolly, J. F. (1987). Psychopathy and event-related brain potentials (ERPs) associated with attention to speech stimuli. *Personality and Individual Differences*, 8, 175-184.
- Kiehl, K. A., Bates, A. T., Laurens, K. R., Hare, R. D., & Liddle, P. F. (2006). Brain potentials implicate temporal lobe abnormalities in criminal psychopaths. *Journal of Abnormal Psychology*, 115, 443-453.
- Kiehl, K. A., Hare, R. D., Liddle, P. F., & McDonald, J. J. (1999). Reduced P300 responses in criminal psychopaths during a visual oddball task. *Biological Psychiatry*, 45, 1498-1507.

## REFERENCES

- Kilpatrick, D. G., Edmunds, C. N., & Seymour, A. K. (1992). *Rape in America: A report to the nation*. Arlington, VA: National Center for Victims of Crime; Charleston, SC: Medical University of South Carolina.
- Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the Control Question Polygraph Technique. *Law and Human Behavior*, 12, 79-90.
- Kok, A., Ramautar, J. R., De Ruiter, M. B., Band, G. P., & Ridderinkhof, K. R. (2004). ERP components associated with successful and unsuccessful stopping in a stop-signal task. *Psychophysiology*, 41, 9-20.
- Kokish, R. (2003). The current role of post-conviction sex offender polygraph testing in sex offender treatment. *Journal of Child Sexual Abuse*, 12, 175-194.
- Kokish, R., Levenson, J. S., & Blasingame, G. D. (2005). Post-conviction sex offender polygraph examination: Client-reported perceptions of utility and accuracy. *Sexual Abuse: A Journal of Research and Treatment*, 17, 211-221.
- Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., & George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, 58, 605-613.
- Kozel, F. A., Padgett, T. M., & George, M. S. (2004). A replication study of the neural correlates of deception. *Behavioral Neuroscience*, 118, 852-856.
- Krapohl, D. J. (2002). The polygraph in personnel screening. In M. Kleiner (Ed.), *Handbook of polygraph testing*. (pp. 217-236). San Diego, CA: Academic Press.
- Krapohl, D. J., Ryan, A. H., & Shull, K. W. (2002). Voice stress devices and the detection of lies. *Polygraph*, 31, 43-48.
- Langleben, D. D., Loughhead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., et al. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, 26, 262-272.
- Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., et al. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *Neuroimage*, 15, 727-732.
- Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior*, 34, 37-59.
- Lapham, S. C., C'De Baca, J., Lapidus, J., & McMillan, G. P. (2007). Randomized sanctions to reduce re-offense among repeat impaired-driving offenders. *Addiction*, 102, 1618-1625.
- Larson, J. A. (1932). *Lying and it's detection: A study of deception and detection tests*. Chicago: University of Chicago Press.
- Lee, T. M., Liu, H. L., Chan, C. C., Ng, Y. B., Fox, P. T., & Gao, J. H. (2005). Neural correlates of feigned memory impairment. *Neuroimage*, 28, 305-313.
- Li, C. S., Huang, C., Constable, R. T., & Sinha, R. (2006). Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. *Journal of Neuroscience*, 26, 186-192.
- Li, C. S., Huang, C., Yan, P., Paliwal, P., Constable, R. T., & Sinha, R. (2008). Neural correlates of posterror slowing during a stop signal task: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*.
- Lieblich, I., Ben Shakhar, G., & Kugelmass, S. (1976). Validity of the guilty knowledge technique in a prisoners' sample. *Journal of Applied Psychology*, 61, 89-93.



## REFERENCES

- Liebllich, I., & Ninio, A. (1972). Detection of suppressed involvement with information through a forced number-guessing technique. *Acta Psychologica*, 36, 381-387.
- Liebllich, I., Shaham, E., & Ninio, A. (1976). Effects of time stress and stimulus-response set size on the efficiency of detection of involvement with suppressed information through the use of the forced number-guessing technique. *Acta Psychologica*, 40, 75-84.
- Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, 66, 488-524.
- Lim, C. L., Gordon, E., Rennie, C., Wright, J. J., Bahramali, H., Li, W. M., et al. (1999). Dynamics of SCR, EEG, and ERP activity in an oddball paradigm with short interstimulus intervals. *Psychophysiology*, 36, 543-551.
- Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: A meta-analysis. *Psychological Bulletin*, 130, 531-552.
- Lösel, F. (2000). The efficacy of sexual offender treatment: A review of German and international evaluations. In P. J. van Koppen & N. Roos (Eds.), *Rationality, information and progress in psychology and law* (pp. 145-170). Maastricht: Metajuridica Publications.
- Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied*, 2, 164-177.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43, 385-388.
- Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, 44, 258-262.
- Lykken, D. T. (1998). *A tremor in the blood*. New York: Plenum Press.
- Lynn, R. (1966). *Attention, Arousal and the Orienting Reaction*. New York: Pergamon.
- MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *Journal of Applied Psychology*, 86, 674-683.
- Malesky, L. A., & Packer, I. K. (2004). Sex offender laws. *Journal of the American Academy of Psychiatry and the Law*, 32, 209-211.
- Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: final results from California's sex offender treatment and evaluation project (SOTEP). *Sex Abuse*, 17, 79-107.
- Marston, W. M. (1938). *The lie detector test*. New York: Richard R. Smith.
- McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 Nationwide Survey*. Brandon, VT: Safer Society Press.
- McGrath, R. J., Cumming, G. F., Hoke, S. E., & Bonn-Miller, M. O. (2007). Outcomes in a community sex offender treatment program: A comparison between polygraphed and matched non-polygraphed offenders. *Sex Abuse*, 19, 381-393.
- Merckelbach, H., Hauer, B., & Rassin, E. (2002). Symptom validity testing of a feigned dissociative amnesia: A simulation study. *Psychology, Crime, and Law*, 8, 311-318.
- Merckelbach, H., & Meijer, E. (2008). Leugendetector helpt niet bij sollicitatie. *NRC*, 12 januari 2008.

## REFERENCES

- Merckelbach, H. L. G. J., & Meijer, E. H. (2007). Trillende stem, bloeiende leugen: over stemanalyse als vorm van leugendetectie. *Neuropsychologie*, 11, 63-67.
- Mertens, R., & Allen, J. J. (2008). The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, 45, 286-298.
- Meyers, J. E., & Diep, A. (2000). Assessment of malingering in chronic pain patients using neuropsychological tests. *Applied Neuropsychology*, 7, 133-139.
- Miller, A. R., & Rosenfeld, J. P. (2004). Response-specific scalp distributions in deception detection and ERP correlates of psychopathic personality traits. *Journal of Psychophysiology*, 18, 13-26.
- Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., et al. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience*, 14, 1200-1214.
- Miyake, Y., Mizutani, M., & Yamahura, T. (1993). Event-related potentials as an indicator of detecting information in field polygraph examinations. *Polygraph*, 22, 131-149.
- Morgan III, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., et al. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry*, 265-279.
- Munsterberg, H. (1908). *On the witness stand*. New York: The McClure Company.
- Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In M. Kleiner (Ed.), *Handbook of polygraph testing*. (pp. 49-86). San Diego, CA: Academic Press.
- National Research Council. (2003). *The polygraph and lie detection. Committee to review the scientific evidence on the polygraph. Division of behavioral and social sciences and education*. Washington, DC: The National Academic Press.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665-675.
- Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: a critical review of past and present strategies. *Archives of Clinical Neuropsychology*, 9, 501-552.
- Ninomiya, H., Onitsuka, T., Chen, C., Sato, E., & Tashiro, N. (1998). P300 in response to the subject's own face. *Psychiatry and Clinical Neurosciences*, 52, 519-522.
- Nunez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage*, 25, 267-277.
- Offe, H. (2001). Fallbericht: Polygraphentest im Familienverfahren. *Praxis der Rechtspsychologie*, 11, 55-57.
- Offe, H., & Offe, S. (2007). The comparison question test: Does it work and if so how? *Law and Human Behavior*, 31, 291-303.
- Office of Technology Assessment. (1983). *Scientific validity of polygraph testing: A research review and evaluation*. Washington, DC: US Government Printing Office.
- Paller, K. A., Bozic, V. S., Ranganath, C., Grabowecky, M., & Yamada, S. (1999). Brain waves following remembered faces index conscious recollection. *Cognitive Brain Research*, 7, 519-531.
- Paller, K. A., Gonsalves, B., Grabowecky, M., Bozic, V. S., & Yamada, S. (2000). Electrophysiological correlates of recollecting faces of known and unknown individuals. *Neuroimage*, 11, 98-110.



## REFERENCES

- Paller, K. A., Ranganath, C., Gonsalves, B., LaBar, K. S., Parrish, T. B., Gitelman, D. R., et al. (2003). Neural correlates of person recognition. *Learning & Memory*, 10, 253-260.
- Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills*, 57, 367-372.
- Piquerez, G. (1987). *Précis de procédure pénale Suisse*. Lausanne: Payot.
- Podlesney, J. A. (1993). Is the guilty knowledge technique applicable in criminal investigations? A review of FBI case records. *Crime Laboratory Digest*, 20, 57-61.
- Podlesney, J. A. (2003). A paucity of operable case facts restricts applicability of the guilty knowledge technique in FBI criminal polygraph examinations. *Forensic Science Communications*, 5, Retrieved April, 16, 2007 from <http://www.fbi.gov/hq/lab/fsc/backissu/july2003/index.htm>.
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10, 59-63.
- Polich, J. (1987). Comparison of P300 from a passive tone sequence paradigm and an active discrimination task. *Psychophysiology*, 24, 41-46.
- Polich, J. (1989). P300 from a passive auditory paradigm. *Electroencephalography and Clinical Neurophysiology*, 74, 312-320.
- Polich, J. (1990). P300, probability, and interstimulus interval. *Psychophysiology*, 27, 396-403.
- Pollatos, O., Gramann, K., & Schandry, R. (2007). Neural systems connecting interoceptive awareness and feelings. *Human Brain Mapping*, 28, 9-18.
- Pollatos, O., Herbert, B. M., Matthias, E., & Schandry, R. (2007). Heart rate response after emotional picture presentation is modulated by interoceptive awareness. *International Journal of Psychophysiology*, 63, 117-124.
- Pollina, D. A., Dollins, A. B., Senter, S. M., Brown, T. E., Pavlidis, I., Levine, J. A., et al. (2006). Facial skin surface temperature changes during a "concealed information" test. *Annals of Biomedical Engineering*, 34, 1182-1189.
- Porter, S., & Yuille, J. C. (1995). Credibility assessment of criminal suspects through statements analysis. *Psychology, Crime and Law*, 1, 319-331.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Raine, A. (1989). Evoked potentials and psychopathy. *International Journal of Psychophysiology*, 8, 1-16.
- Raine, A., & Venables, P. H. (1988). Enhanced P3 evoked potentials and longer P3 recovery times in psychopaths. *Psychophysiology*, 25, 30-38.
- Ramautar, J. R., Kok, A., & Ridderinkhof, K. R. (2004). Effects of stop-signal probability in the stop-signal paradigm: The N2/P3 complex further validated. *Brain and Cognition*, 56, 234-252.
- Raskin, D. C., & Honts, C. R. (2002). The Comparison Question Test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. (pp. 1-47). San Diego, CA: Academic Press.
- Read, W., Nenov, V. I., & Halgren, E. (1994). Role of inhibition in memory retrieval by hippocampal area CA3. *Neuroscience and Biobehavioral Reviews*, 18, 55-68.
- Reid, J. E. (1947). A revised questioning technique in liedetection tests. *Journal of Criminal Law and Criminology*, 37, 542-547.
- Renault, B., Signoret, J. L., Debrulle, B., Breton, F., & Bolgert, F. (1989). Brain potentials reveal covert facial recognition in prosopagnosia. *Neuropsychologia*, 27, 905-912.
- Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and analysis. *Psychological Bulletin*, 114, 363-375.

## REFERENCES

- Rosen, G. M., & Powel, J. E. (2003). Use of a Symptom Validity Test in the forensic assessment of Posttraumatic Stress Disorder. *Anxiety Disorders, 17*, 361-367.
- Rosenfeld, J. P. (2002). Event-related potentials in the detection of deception, malingering, and false memories. In M. Kleiner (Ed.), *Handbook of polygraph testing*. (pp. 265-286). San Diego, CA: Academic Press.
- Rosenfeld, J. P., Biroshak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology, 60*, 251-259.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdacz, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience, 42*, 157-161.
- Rosenfeld, J. P., Ellwanger, J., & Sweet, J. (1995). Detecting simulated amnesia with event-related brain potentials. *International Journal of Psychophysiology, 19*, 1-11.
- Rosenfeld, J. P., Ellwanger, J. W., Nolan, K., Wu, S., Bermann, R. G., & Sweet, J. (1999). P300 scalp amplitude distribution as an index of deception in a simulated cognitive deficit model. *International Journal of Psychophysiology, 33*, 3-19.
- Rosenfeld, J. P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biological Psychology, 74*, 396-404.
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology, 41*, 205-219.
- Ruff, C. C., Woodward, T. S., Laurens, K. R., & Liddle, P. F. (2001). The role of the anterior cingulate cortex in conflict processing: Evidence from reverse stroop interference. *Neuroimage, 14*, 1150-1158.
- Salisbury, D. F., Griggs, C. B., Shenton, M. E., & McCarley, R. W. (2004). The NoGo P300 'anteriorization' effect and response inhibition. *Clinical Neurophysiology, 115*, 1550-1558.
- Salisbury, D. F., Rutherford, B., Shenton, M. E., & McCarley, R. W. (2001). Button-pressing affects P300 amplitude and scalp topography. *Clinical Neurophysiology, 112*, 1676-1684.
- Schmid, N. (1993). *Strafprozessrecht: Eine Einführung auf der Grundlage des Strafprozessrechtes des Kantons Zürich und des Bundes*. Zürich: Schulthess.
- Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology, 37*, 257-261.
- Schwartz, & Cellini. (1995). *The Sex Offender: Corrections, Treatment and Legal Practice*. Kingston, NJ: Civic Research Institute.
- Schweinberger, S. R., & Burton, A. M. (2003). Covert recognition and the neural system for face processing. *Cortex, 39*, 9-30.
- Schweinberger, S. R., Pickering, E. C., Burton, A. M., & Kaufmann, J. M. (2002). Human brain potential correlates of repetition priming in face and name recognition. *Neuropsychologia, 40*, 2057-2073.
- Schweitzer, R., & Dwyer, J. (2003). Sex crime recidivism: Evaluation of a sexual offender treatment program. *Journal of Interpersonal Violence, 18*, 1292-1310.
- Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement, 23*, 153-158.



## REFERENCES

- Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, 23, 695-703.
- Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge". *Journal of Applied Psychology*, 85, 30-37.
- Siegel, S. (1956). *Nonparametric statistics*. New York: McGraw-Hill.
- Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends in Cognitive Sciences*, 12, 48-53.
- Smith, N. (2001). *Reading between the lines: An evaluation of the scientific content analysis technique (SCAN)*. London: UK Home Office, Research, Development and Statistics Directorate.
- Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Reviews in Neurobiology*, 14, 199-224.
- Sommer, W., Leuthold, H., & Matt, J. (1998). The expectancies that govern the P300 amplitude are mostly automatic and unconscious. *Behavioral and Brain Sciences*, 21, 149-150.
- Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: Complex vs. simple paradigms. *International Journal of Psychophysiology*, 40, 173-180.
- Spence, S. A. (2008). Playing devil's advocate: The case against fMRI lie detection. *Legal and Criminological Psychology*, 13, 11-26.
- Spence, S. A., Farrow, T. F., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, 12, 2849-2853.
- Spence, S. A., Hunter, M. D., Farrow, T. F., Green, R. D., Leung, D. H., Hughes, C. J., et al. (2004). A cognitive neurobiological account of deception: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 359, 1755-1762.
- Spence, S. A., Kaylor-Hughes, C. J., Brook, M. L., Lankappa, S. T., & Wilkinson, I. D. (in press). 'Munchausen's syndrome by proxy' or a 'miscarriage of justice'? An initial application of functional neuroimaging to the question of guilt versus innocence. *European Psychiatry*.
- Steller, M., & Dahle, K. P. (1999). Wissenschaftliches Gutachten: Grundlagen, Methoden und Anwendungs-probleme psychophysiologischer Aussage- bzw. Täterschaftsbeurteilung ("Polygraphie", "Lügendetektion"). *Praxis der Rechtspsychologie*, 9, 127-204.
- Stern, J. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording*. New York: Oxford University Press.
- Stroemwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practicionars' beliefs about deception. In L. A. Stroemwall & P. A. Granhag (Eds.), *The detection of deception in forensic contexts* (pp. 229-250). Camebridge: Camebridge University Press.
- Strong, J. W. (Ed.). (1999). *McCormick on evidence*. St. Paul, MI: West.
- Suchday, S., & Larkin, K. T. (2004). Psychophysiological responses to anger provocation among Asian Indian and White men. *International Journal of Behavioral Medicine*, 11, 71-80.
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150, 1187-1188.

## REFERENCES

- Timm, H. W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *Journal of Applied Psychology*, 67, 391-400.
- Traest, P. (2001). Recente tendensen in het bewijsrecht in strafzaken. In Vlaamse Conferentie der Balie van Gent (Ed.), *Straf recht? Strafrecht*. Antwerpen: Maklu.
- Tysse, J. E. (2005). Note: The right to an "imperfect" trial - amnesia, malingering, and competency to stand trial. *William Mitchell Law Review*, 32, 353-387.
- Undeutsch, U. (1996). Die Untersuchung mit dem Polygraphen (lügendetektor) - eine wissenschaftliche Methode zum Nachweis der Unschuld. *Zeitschrift für das gesamte Familienrecht*, 6, 329-331.
- Undeutsch, U., & Klein, G. (1999). Wissenschaftliches Gutachten zum Beweiswert physiopsychologischer Untersuchungen. *Praxis der Rechtspsychologie*, 9, 45-126.
- van Koppen, P. J., Boelhouwer, A. J. W., Merckelbach, H., & Verbaten, M. N. (1996). *Leugendetecie in actie: Het gebruik van de polygraaf in de praktijk*. Leiden: Nederland: Studiecentrum Criminaliteit en Rechtshandhaving (NSCR) (Rapport aangeboden aan de minister van Justitie).
- Verhaegen, P. (2000). *De polygraaf en zijn potentiële plaats in het Belgisch strafrechtelijk bewijsstelsel: Advies voor de heer Minister van Justitie en het College van Procureurs-generaal inzake het gebruik van de polygraaf in de Belgische strafrechtsprocedure*. Brussel: Ministerie van Justitie, Dienst voor Strafrechtelijk Beleid.
- Verschuere, B., Crombez, G., De Clercq, A., & Koster, E. H. (2004). Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology*, 41, 461-466.
- Verschuere, B., Crombez, G., Declercq, A., & Koster, E. H. W. (2005). Psychopathic traits and autonomic responding to concealed information in a prison sample. *Psychophysiology*, 42, 239-245.
- Verschuere, B., Crombez, G., Koster, E. H., & De Clercq, A. (2007). Antisociality, underarousal and the validity of the Concealed Information Polygraph Test. *Biological Psychology*, 74, 309-318.
- Verschuere, B., Crombez, G., Koster, E. H., Van Bockstaele, B., & De Clercq, A. (2007). Startling secrets: Startle eye blink modulation by concealed crime information. *Biological Psychology*, 76, 52-60.
- Verschuere, B., Meijer, E., & Crombez, G. (in press). Symptom Validity Testing for the detection of simulated amnesia. Not robust to coaching. *Psychology, Crime and Law*.
- Volokh, A. (1997). n Guilty men. *University of Pennsylvania Law Review*, 146, 173-211.
- Vrij, A. (2008). Beliefs about nonverbal and verbal cues to deception. In A. Vrij (Ed.), *Detecting lies and deceit* (pp. 115-140). Chirchester: Wiley.
- Vrij, A. (2008). *Detecting lies and deceit. Pitfalls and opportunities*. Chichester: Wiley.
- Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin*, 77, 65-72.
- Ward, T., Hudson, S. M., Johnston, L., & Marshall, W. L. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review*, 17, 479-507.
- Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, 26, 208-221.
- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, 20, 470-477.
- Werkgroep Leugendetecie. (1993). *De leugendetector: Een verkennend onderzoek naar de mogelijkheden van de polygraaf*. Den Haag: Ministerie van Justitie.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.



## REFERENCES

- Wiens, S., Mezzacappa, E. S., & Katkin, E. (2000). Heartbeat detection and the experience of emotions. *Cognition and Emotion*, 14, 417-427.
- Wiggins, E. C., & Brandt, J. (1988). The detection of simulated amnesia. *Law and Human Behavior*, 12, 57-78.
- Wilcox, D. T., & Sosnowski, D. E. (2005). Polygraph examination of British sexual offenders: A pilot study on sexual history disclosure testing. *Journal of Sexual Aggression*, 11, 3-23.
- Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research*, 2, 149-168.
- Woestenburg, J. C., Verbaten, M. N., & Slangen, J. L. (1981). The influence of information on habituation of the "Wiener" filtered visual event related potential and the skin conductance reaction. *Biological Psychology*, 13, 189-201.
- Woestenburg, J. C., Verbaten, M. N., & Slangen, J. L. (1983). Stimulus information and habituation of the visual event related potential and the skin conductance reaction under task-relevance conditions. *Biological Psychology*, 16, 225-240.
- Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research*, 1100, 125-135.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441-517.

## Summaries

### *English*

This dissertation deals with the use of polygraph tests for the detection of deception. The first two chapters deal with the growing use of the Control Question polygraph Test (CQT) in Europe. In chapter 1, the nature of this test is explained, and we review how far it has penetrated into the legal arena of different European countries. The CQT is used in police investigations on a relatively large scale in countries like Belgium and Finland. Courts around Europe seem to be aware of the controversial nature of this polygraph technique, and have generally rejected it as evidence. Due to the free evidence system, however, this has never been a general ban. Moreover, tests can be used in criminal investigations, without ever reaching the courtroom.

Chapter 2 discusses a specific application of the CQT that is also becoming increasingly popular in Europe, namely its use in the treatment and risk assessment of sex offenders. We critically examine the validity and utility of this application. Our review shows that there is no evidence for the accuracy of the CQT in this application, and that the results in the field can largely be explained by the bogus pipeline phenomenon. Furthermore, the available evidence for the claims about the clinical potential of polygraph tests is weak, if not absent. We conclude that portraying post-conviction polygraph testing as analogous to urine analysis in the treatment of drug addicts, as is done by American clinicians, is inaccurate, misleading, and ultimately, risky.

The remainder of this dissertation deals with the use of the Concealed Information polygraph Test (CIT). Chapter 3 explains how this test can be used to determine the presence or absence of guilty knowledge, and the dependent measures that are used with this test (Skin Conductance Responses (SCRs), the P300 component of the event-related potential, and a forced choice procedure) are discussed.

Chapter 4 addressed the use of the P300 to detect concealed face recognition. The results of two experiments show that detection of concealed face recognition is highly successful when stimuli depict persons who are personally highly familiar, and instructions to conceal recognition are given. When pictures depict recognized, but personally less familiar faces, and no specific instructions to conceal recognition are given, detection is unsuccessful. These findings indicate that pictures of faces can



## SUMMARIES

be used in a P300 based CIT, and that mere recognition is not sufficient for successful detection of concealed information.

Chapter 5 examines the association between psychopathic traits, SCR, and the P300 in a CIT. Participants performed a mock crime, after which they were subjected to a CIT. During this CIT, skin conductance and EEG were measured simultaneously, while stimuli were presented with a short inter stimulus interval (ISI). To overcome the problem of overlapping SCRs with these ISIs, stimuli were presented in a specially balanced order. Results showed that detection of concealed information with P300 recovery, but not with P300 amplitude nor with SCR correlated with psychopathic traits. Furthermore, SCR did not correlate with P300. To the extent that psychopathy moderates autonomic responding in a CIT thereby reducing its diagnostic accuracy, measurement of the P300 may be a solution.

In chapter 6, we tested whether the sensitivity of a CIT procedure could be increased by adding an independent measure that is based on an entirely different psychological mechanism than the psychophysiological measures. In two experiments, we explored whether the accuracy of a CIT procedure could be increased by adding SVT, a relatively simple, forced-choice, self-report procedure that has previously been used to detect malingering in various contexts. Results of a feigned amnesia experiment, but not from a mock crime experiment showed that a combination measure of both tests yielded better detection than either test alone.

In the final chapter, theoretical and practical implications are discussed. Attention is given to recent developments in the domain of lie detection, and directions for future research are discussed. It is concluded that the CIT will be a valuable addition to the investigative authorities' toolkit.

***Nederlands***

Dit proefschrift gaat over leugendetectie met behulp van de polygraaf. De eerste twee hoofdstukken gaan over het toenemende gebruik van de controle vragen variant van de polygraaf in Europa. In hoofdstuk 1 word besproken hoe deze test in zijn werk gaat, en bespreken we in hoeverre de test in de rechtssystemen van de verschillende Europese landen is doorgedrongen. De controle vragen variant wordt in landen als België en Finland op relatief grote schaal ingezet. Rechtbanken lijken zich bewust van de omstreden status ervan, en de uitslag ervan wordt dan ook niet als bewijs geaccepteerd. Omdat er in veel Europese landen vrije bewijsleer geldt heeft dit echter nooit in een algeheel verbod geresulteerd. De test kan ook gebruikt worden in de politionele opsporing, zonder ooit de rechtbank te halen.

Hoofdstuk 2 behandelt een specifieke toepassing van de controle vragen polygraaf test die zich ook in Nederland in een toegenomen belangstelling mag verheugen, namelijk het periodiek testen van zedendelinquenten. Het idee achter deze methode is dat men met behulp van de polygraaf gedrag (bijvoorbeeld risicogedrag tijdens een proefverlof) beter in kaart kan brengen dan met de gebruikelijke zelfrapportage. Uit de evaluatie van de beschikbare wetenschappelijke literatuur blijkt echter dat evidentie voor de nauwkeurigheid van de controle vragen polygraaf test in deze toepassing ontbreekt, en de in de praktijk behaalde resultaten kunnen ook aan de hand van de intimiderende werking van de polygraaf verklaard worden. Hiernaast ontbreekt evidentie voor het effect van het gebruik van de polygraaf op recidive. Allerhande optimistische claims van behandelaars over het nut van de polygraaf bij het behandelen van zedendelinquenten zijn dan ook prematuur, en het meewegen van de uitslag van een polygraaf test bij de besluitvorming over voorwaardelijke invrijheidstelling en vrijheidsbeneming van veroordeelden kan tot ernstige fouten leiden.

Deel twee van dit proefschrift gaat over de *Concealed Information* variant van de polygraaf. Bij deze variant wordt de polygraaf gebruikt om de aan of afwezigheid van daderkennis aan te tonen. Centraal in dit deel staat de vraag hoe de nauwkeurigheid van deze test verbeterd kan worden. In hoofdstuk drie wordt deze test beschreven, en passeren de afhankelijke maten die in deze dissertatie aan bod komen (huidgeleiding, P300 hersengolf en *forced choice*) de revue.

Hoofdstuk vier beantwoordt de vraag of de P300 hersengolf gebruikt kan worden om bekendheid met gezichten aan te tonen. De resultaten van twee experimenten laten zien dat kan zolang het gaat om zeer bekende



## SUMMARIES

gezichten (broer/zus, goede vriend/vriendin), en er instructies om herkenning te verbergen worden gegeven. Als het gaat om herkende, maar minder bekende gezichten, in de afwezigheid van instructies om herkenning te verbergen, bleef de P300 uit. Deze resultaten laten zien dat herkenning alleen onvoldoende is voor een succesvolle, op de P300 gebaseerde *Concealed Information* variant van leugendetectie.

In hoofdstuk vijf staat de relatie tussen psychopatische trekken, huidgeleiding en de P300 in een *Concealed Information* test centraal. Hersenactiviteit en huidgeleiding werden gelijktijdig geregistreerd terwijl stimuli kort na elkaar werden aangeboden. Om de overlappende huidgeleidingsresponsen van elkaar te scheiden werden de stimuli in een speciaal gebalanceerde volgorde aangeboden. P300 *recovery*, correleerde positief met psychopatische trekken. P300 amplitude en huidgeleiding correleerden niet met psychopatische trekken. Huidgeleiding en P300 waren ook ongecorreleerd. Dit betekent dat het meten van de P300 een oplossing kan zijn voor de verminderde autonome responsiviteit die met psychopathie gepaard gaat.

In hoofdstuk zes onderzochten we of de sensitiviteit van een *Concealed Information* test verbeterd kan worden door een maat toe te voegen die op een ander psychologisch mechanisme berust dan de psychofysiologische maten. Het gaat hier om een *forced choice* procedure die eerder gebruikt is om verschillende vormen van *malingering* te detecteren. Als de proefpersonen werd gevraagd om amnesie voor hun identiteit te veinzen leidde een combinatie van huidgeleiding en *forced choice* tot een hogere sensitiviteit dan beide maten apart. Dit was echter niet het geval in een experiment waar de proefpersonen werd gevraagd om een nepmisdrijf te plegen.

In het laatste hoofdstuk worden de theoretische en praktische implicaties van het onderzoek besproken. Verder wordt er aandacht besteed aan recente ontwikkelingen in het veld, met als eindconclusie dat de *Concealed Information* variant een waardevolle aanvulling op het arsenaal van politionele opsporingsmiddelen vormt.

## CURRICULUM VITAE

### **Curriculum vitae**

Ewout Meijer was born in Apeldoorn on January 2nd 1975. He received his VWO degree from *sonsbeek college* in 1994. That year he started studying Health Sciences at Maastricht University. During this study, he spent a year at the kinesiology Department of the University of Maryland at College Park (USA), and graduated in both Movement Sciences and Mental Health Sciences in 2003. He currently holds a position as post doctoral researcher at the department of Psychology and Neuroscience of Maastricht University.



## PUBLICATIONS

### Publications

#### In English

- Verschuere, B., **Meijer, E.**, & Crombez, G. (in press). Symptom Validity Testing for the detection of simulated amnesia. Not robust to coaching. *Psychology, Crime and Law*.
- Meijer, E.H.**, Verschuere, B., Merckelbach, H.L.G.J., & Crombez, G. (in press). Sex offender management using the polygraph: A critical review. *International Journal of Law and Psychiatry*.
- Verschuere, B., **Meijer, E.**, & Merckelbach, H. (2008). The Quadri-Track Zone Comparison Technique: it's just not science. A critique to Mangan, Armitage, and Adams (2008). *Physiology and behavior*, 95, 27-28.
- Smulders, F.T.Y., & **Meijer, E.H.** (2008). Extraversion and performance: A cognitive-energetical approach. *Personality and Individual Differences*, 44, 475-486.
- Meijer, E.H.**, Smulders, F.T.Y., Merckelbach, H.L.G.J., & Wolf, A.G. (2007). The P300 is sensitive to concealed face recognition. *International Journal of Psychophysiology*, 66, 231-237.
- Meijer, E.H.**, Smulders, F.T.Y., Johnston, J.E., & Merckelbach, H.L.G.J. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology*, 44, 814-822.

#### In Dutch

- Meijer, E.H.**, & Merckelbach, H. (2008). Leugendetectie: oude waarheden en nieuwe technologieën. *Justitiële Verkenningen*, 34, 42-53.
- Merckelbach, H.L.G.J., & **Meijer, E.H.** (2007). Trillende stem, bloeiende leugen: over stemanalyse als vorm van leugendetectie. *Neuropraxis*, 11, 63-67.
- Meijer, E.H.**, & Merckelbach, H.L.G.J. (2006). De leugendetector bij zedendelinquenten: niet doen! *Nederlands Juristenblad*, 79-84.
- Meijer, E.H.**, & Merckelbach, H.L.G.J. (2004). Leugendetectie in twee bedrijven. *Algemeen Politieblad*, 11, 12-14.
- Meijer, E.H.** (2004). Leugendetectie en wetenschap: vriend of vijand. *De Psycholoog*, 39, 174-179.
- Merckelbach, H.L.G.J., & **Meijer, E.H.** (2004). De geheugendetector. Mogelijkheden en beperkingen. *Justitiële Verkenningen*, 30, 62-78.

## PUBLICATIONS

### Book Chapters

- Meijer, E. H., & van Koppen, P. J.** (2008). Lie detectors and the law: The use of the polygraph in Europe. In D. Canter & R. Zukauskienė (Eds.), *Psychology and law: Bridging the gap* (pp. 31-50). Aldershot, UK: Ashgate Publishing.
- van Oorsouw, K., & **Meijer, E.** (2008). Symptom Validity Testing. In T. Giesbrecht, C. de Ruiter & M. Jelicic (Eds.), *Forensisch psychodiagnostisch gereedschap. Malingering, psychopathie en andere persoonlijkheidstreken*. Amsterdam: Harcourt Assessment BV.

### Miscellaneous

- Merckelbach, H., & **Meijer, E.** (2008). Leugendetector helpt niet bij sollicitatie. *NRC*, 12 januari 2008.
- Meijer, E., & Merckelbach, H.** (2007). Leugendetector geeft vals gevoel van veiligheid. *Volkskrant*, 21 augustus 2007.



## DANKWOORD

### **Dankwoord**

Voor de tot standkoming van dit proefschrift ben ik een aantal personen dank verschuldigd. Als eerste gaat noem ik Harald en Fren. Voor de kans om te promoveren, hun inhoudelijke begeleiding daarbij, maar zeker ook voor de vrijheid waarmee ik het project vorm heb mogen geven.

Ook mijn (ex)collega's ben ik dank verschuldigd: Beatrijs, Corine, Elke, Henry, Ingrid Jill, Kim, Linsey, Maarten, Maartje, Marko, Saskia en Tom, maar ook de overige collega's bij de FdP en FdR. Zonder jullie was het maar half zo leuk geweest.

Mijn onderzoek beschreven in deze dissertatie was niet mogelijk geweest zonder de bijdrage van een aantal studenten. Jim Johnston, Ann Wolf, Marleen van Osch en Nieke Elbers, bedankt.

Stefan en Timo, ik ben blij dat jullie mijn paranimfen willen zijn.

Tot slot mogen natuurlijk mijn ouders in dit rijtje niet ontbreken. Bedankt dat jullie altijd achter mij hebben gestaan.